

THIS WEEK

EDITORIALS

POLICY Obama must let the legal ruling on contraception stand **p.138**



WORLD VIEW Clive Hamilton argues against geoengineering research **p.139**

PSYCHOLOGY Gaze bias explains why the eyes always have it **p.140**

The right to speak out

Controversy over the results touted by a genetic-ancestry firm has highlighted the need for reform of the United Kingdom's restrictive libel law.

What do Adam, Eve and the Queen of Sheba have to do with libel reform? Ask David Balding and Mark Thomas, geneticists at University College London (UCL) who received legal threats after they criticized the claims of a firm that sells people details of their genetic ancestry. Or ask the student journalists who feared a libel lawsuit if they covered the row in their university newspaper. Or the senate committee of the same university that was forced to slap down its own rector for actions contrary to academic freedom.

It is a messy and perhaps uniquely British farce, and one that highlights the desperate need to change English libel laws. And it shows why long-promised reform, due to be discussed again in Parliament later this month, might not go far enough.

The story began last July, when Balding and Thomas heard Alistair Moffat, chief executive of genetic-analysis company BritainsDNA and rector of the University of St Andrews, tell BBC radio that his firm had discovered Eve's grandson and nine Britons directly descended from the Queen of Sheba. He added that a volcanic eruption 70,000 years ago had wiped out all human lineages except those of Adam and Eve, that 97% of men with the last name of Cohen share a single genetic marker and that one-third of British men are descended from the country's founding lineages. "The Bible, through BritainsDNA, is really starting to come alive," announced Moffat, a historian.

Nonsense, Balding and Thomas said in a series of private letters and e-mails to Moffat and his company. Noting that scientists at BritainsDNA had trained at UCL, and concerned for the university's reputation, Balding wrote: "Do you hope to maintain respectable scientific careers and also lend credibility to such nonsense?" Last September, he and Thomas received a letter from Moffat's lawyers, demanding that they did "not report or state as a matter of undisputed fact that our client's science is 'wrong' or untrue".

As *Nature* knows to its cost, libel cases are expensive and time-consuming. But Balding and Thomas decided that they were on solid enough scientific ground to go public with their criticism of the science.

BritainsDNA looks at 'uni-parental' genetic markers on the maternally transmitted mitochondria and the paternally inherited Y chromosome. The Adam and Eve to whom Moffat referred are not those in the Bible, but individuals who lived in Africa 150,000–200,000 years ago and who had Y chromosomes and mitochondria that can be traced to humans alive today. These individuals are not special: their lineages are two among many that contributed DNA to modern humans.

Moffat's claims about Adam and Eve, then, were meaningless, Balding and Thomas argued in several blog posts. Others, they considered, were flat wrong. Nobody knows which genetic markers the Queen of Sheba possessed, they said, and it is very unlikely that the eruption of Mount Toba in Indonesia 70,000 years ago wiped out all genetic lineages but those of 'Adam' and 'Eve'.

Such overreach is a worldwide problem. The American Society of Human Genetics in Bethesda, Maryland, has twice issued policy

papers that call on genetic-ancestry companies to state clearly the limitations and uncertainties in the information that they sell.

Balding and Thomas never heard back from Moffat or his solicitors. But a student journalist at the University of St Andrews, where Moffat's position of rector is non-executive, received legal threats when he attempted to report the duo's concerns. Jonathan Bucks, news editor for *The Saint*, says that Moffat repeatedly warned that if necessary, he would take legal action over anything the student newspaper published.

Senior members of the St Andrews academic senate launched an investigation into Moffat's threats to Balding and Thomas, and concluded that they were "contrary to the principles of academic freedom and scientific debate in a matter of public interest". The student newspaper ran its story, and did an excellent job.

Moffat may perhaps have been bluffing about his legal threats (he and his lawyers did not respond to queries from *Nature*) but under present laws, just the threat of a lawsuit is often enough to silence reasonable scientific criticism. It is right that reforms currently being considered by the nation's Parliament would offer special protection to claims made in peer-reviewed articles. But only a clear public-interest defence for all scientific debate, at present not included in the reformed laws, will grant scientists the freedom that they need to speak out to safeguard public trust. ■

Energy crossroads

Everyone should wish Germany well in its great experiment in renewable energy.

The daughter of an East German pastor, the German chancellor Angela Merkel owes her political career to the peaceful revolution of autumn 1989 — *die Wende* (the turnaround) — that led to Germany's reunification. And she may well be remembered for another turning point in the nation's post-war history — *die Energiewende*.

Merkel has long been a driving force behind efforts to tackle global warming. The *Energiewende*, Germany's solo effort to radically shift its economy to one dominated by renewable sources of energy, is a pragmatic alternative to the largely fruitless efforts of international climate-change diplomacy. If the country's do-it-yourself tactics bear fruit, it would be cause for renewed optimism in the fight against anthropogenic climate change.

As we report on page 156, Germany aims by mid-century to produce at least 80% of its electricity, and more than half of the energy it needs for heating, from renewable energy sources such as sunlight and wind. Thousands of scientists and engineers are developing the power-storage and transmission technologies required to accommodate tomorrow's fluctuating energy sources. The task is enormous. Nevertheless, given the growth rate of renewable energies in recent years — and the high level of public acceptance of the policy — the goals could be achieved even earlier than planned.

In the public's perception, the *Energiewende* has less to do with global warming than it has with the nuclear disaster at Fukushima in Japan two years ago, which accelerated the proposed phase-out of nuclear power in Germany. In fact, the closure of eight nuclear power plants has made Germany — at least temporarily — more dependent on coal, which will make near-term reduction targets for greenhouse-gas emissions harder to achieve.

The German effort is no less impressive for all that. Merkel and the German public are displaying admirable courage in turning the country into a laboratory for energy policies (and technologies) which could become models for many industrialized countries. That the nation is running this costly and risky self-experiment against the background of a European arch financial crisis should stand as an enduring testament to Merkel's leadership.

Germany, with its population of some 80 million, is an ideal test ground. The country can afford the *Energiewende* — which some commentators estimate will cost more than €1 trillion (US\$1.3 trillion) — because its economy is doing well. And the country is large and diverse enough — economically, geographically and socially — to make the outcome of the great experiment relevant to the rest of the world.

If German's transition to cleaner energy succeeds, then the country will have learned scientific, technical and economic lessons that it will take to the market place and that will solidify its leadership in green technologies. Countries that might eye the German plans with some scepticism now could eventually build on these technologies when they start to reshape their own energy systems.

If the *Energiewende* founders, however, it will send out a very negative message. Sceptics worldwide will argue that if Germany can't make it work, then nobody can. It is crucial, therefore, that Germany maintains its chosen path through whatever storms may come and even if moans about high electricity prices become more audible.

Ultimately, any truly green economy must include all economic sectors. The *Energiewende* will not be complete without a new approach to transport. It took generous incentives to convince millions of German homeowners to invest in expensive (and aesthetically debatable) rooftop solar panels. Convincing Germans — or any other nation — to switch to battery-driven or electric cars will be even harder. Without incentives for car makers to produce those cars, and for motorists to buy them, it will not happen. But as seen in the rush for the 'cash for clunkers' bonus (as part of the economic stimulus of 2009 the German government paid car owners a €2,500 premium to exchange a car more than nine years old for a new one), such incentives can mobilize massive behavioural change. An effective green-car programme, backed by investment in the necessary infrastructure, would be an unmistakable signpost on the road to the post-fossil-fuel age — and not only in Germany. ■

"Merkel and the German public are displaying admirable courage in turning the country into a laboratory for energy policies."

Time for plan B

A court ruling to remove age limits on access to emergency contraception must prevail.

A federal judge last week ordered the US Food and Drug Administration (FDA) to make the emergency contraceptive Plan B One-Step (levonorgestrel) available over-the-counter to all women and girls, irrespective of age. The ruling is a welcome one, and the administration of President Barack Obama would be wise to end its history of political interference in a scientific issue by declining to appeal the decision.

FDA officials have restricted access to the 'morning after' pill ever since the administration of George W. Bush, despite the consistent opinions of their scientists and advisers, who have been recommending universal over-the-counter access since 2003. The drug, a hormone used in many birth-control pills, has been widely shown to be safe and efficacious, and young teens have proven at least as adept as older women in following instructions for taking it. Nonetheless, the FDA continued to require girls younger than 17 to get a doctor's prescription, and women older than that to follow a burdensome proof-of-age procedure.

The drug regulator's reticence can be traced to pressure from conservatives, who argued — despite studies showing otherwise — that making Plan B easier to access would encourage promiscuity and sexual risk-taking in young teens. Abortion politics also had a role: although the pill is thought to act by preventing or delaying ovulation, it might also prevent implantation of a fertilized zygote.

In 2001, women's and reproductive-rights advocates petitioned the agency to remove the age restriction, without success. Their hope surged when a newly elected Obama publicly committed to letting agency scientists do their jobs, "listening to what they tell us, even when it's inconvenient — especially when it's inconvenient"

(see *Nature* 480, 413; 2011). Then, in December 2011, as FDA head Margaret Hamburg prepared to make the drug available without a prescription to girls younger than 17, her boss overruled her. In her unprecedented action, Kathleen Sebelius, the US secretary of health and human services, speciously noted that 10% of 11-year-old girls are capable of conceiving and implied that young girls could therefore be harmed by the drug. In fact, the only documented adverse effects are nausea and delayed menses, and an 11-year-old can buy a lethal dose of paracetamol for considerably less money and with no restrictions. All the same, Obama backed up Sebelius.

Now, Edward Korman, a senior judge in the eastern district of New York, has called Sebelius's action, which was taken less than a year before a closely contested presidential election, "politically motivated, scientifically unjustified, and contrary to agency precedent". In a scathing decision in a lawsuit against the FDA brought by women and reproductive-rights groups, Korman noted that Plan B One-Step would be one of the safest over-the-counter drugs. He added that the administration's position "is not about the potential misuse of Plan B by 11-year-olds ... [but] an excuse to deprive the overwhelming majority of women of their right to obtain contraceptives without unjustified and burdensome restrictions".

Those restrictions are particularly intrusive because time is of the essence in using the drug: the sooner after intercourse that the tablet is taken, the more likely it is to prevent pregnancy. The need to find an open pharmacy and show proof of age is not trivial, especially for poorer, younger and less-educated women — never mind asking those under 17 to get a doctor's prescription.

The judge's pointed ruling has given the Obama administration another chance to live up to the president's promise to respect agency scientists' findings "especially when it's inconvenient". But if the Department of Justice appeals, and prevails, then the ramifications will extend well beyond the fate of an emergency contraceptive. It will set a troubling precedent for political interference in drug-approval decisions. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xbhunq



No, we should not just 'at least do the research'

The idea of applying geoengineering research to mitigate climate change has not been thought through, argues Clive Hamilton.

Fresh concerns about using geoengineering projects to cool the planet emerged late last month, when scientists at the UK Met Office said that possible unintended consequences demanded global oversight of such schemes. With the Intergovernmental Panel on Climate Change soon to report on it, geoengineering will not go away and neither will the controversy it provokes.

One constant is the call that 'we should at least do the research' so that we can be prepared. In truth, this simple injunction is part of the problem. It rests on a string of questionable assumptions and a naive understanding of the world that owes more to the quaint ideal of the white-coated scientist dispassionately going about the process of knowledge generation than it does to reality.

There are some hard questions for those who believe that 'we should at least do the research'. To start with, who is this 'we'? Is it the 'rogue' geoengineer Russ George, who wants to fertilize the oceans with iron so that he can generate carbon credits to sell? Is it the eccentric Russian Yuri Izrael, who is experimenting with aerosol spraying? How about oil giants such as Exxon-Mobil, which for years funded climate-science disinformation and is now talking up the prospects of geoengineering. Does 'we' mean the Chinese or US military, the organizations with the best access to the equipment needed to deploy a sulphate aerosol shield?

And who should pay for the research? Should it be the public, through national research programmes? Or is it all right for it to be billionaires, backyard tinkerers and oil companies? Shell now funds research into liming the oceans through the Cquestrate project, and ConocoPhillips, among others, is investing in biochar research.

Should the research be transparent, or should it be secret? If we believe that it is too important to be carried out in secret, should corporations be forced to open up their labs? Should the United Nations have weather-technology teams that are empowered to inspect military research facilities in, say, China, in the same way that it sent weapons-inspection teams into Iran? We also need to decide who should oversee and regulate the research. Should it be left to the scientists to regulate themselves, as they proposed at the Asilomar geoengineering conference in 2010? What if those scientists are employed by oil companies or the governments of coal-dependent states?

Should national governments impose ethical standards? Should a UN agency be established to oversee all global research? Should vulnerable nations have a greater say in where research funds are spent and how research is governed, or should their poverty exclude them from any say in potentially world-saving technologies?

Who should own the results of the research?

Should private patents be issued so that individuals or corporations are in a position to decide to sell their planet-saving technologies?

Such fears are not new. US President Dwight Eisenhower warned in 1961 of the "danger that public policy could itself become the captive of a scientific-technological elite". So how do we prevent the formation of a powerful constituency of scientists, investors and politicians after a quick fix, a lobby that could manipulate the political system to downplay or override serious concerns about safety in order to see its technology deployed? And if we do the research and obtain the hoped-for results, and the demands for deployment become overwhelming, who will control what is deployed, and when and where? If deploying a solar shield has divergent effects on precipitation in rich and poor nations, who decides where the rain should fall?

We should have satisfactory answers to these questions before developing the means to engineer the climate. We cannot be content with the idea that the world will somehow muddle through. Research around the world is gathering pace and is answering these questions by default. Frequently, the answers are unpalatable — for example, a slew of private patents has already been taken out over methods to engineer the world's climate. So the call to 'do the research' entrenches a situation in which geoengineering is often carried out by the wrong people, for the wrong reasons and with no oversight, and in the process is creating a lobby group that is likely to press for deployment because it is in its financial or

professional interests to do so.

Political context matters. These concerns are amplified when we think about how we got here. We are entertaining the idea of intervening in the climate system to prevent climatic disaster because of the inability of our political and social systems to implement sharp reductions in greenhouse-gas emissions. If weak political leadership, the power of the fossil-fuel lobby, pervasive wishful thinking and a culture of denial have undermined plan A, why would we not expect those same conditions to govern the development and deployment of plan B?

The 'we should at least do the research' lobby assumes that, if geoengineering research succeeds and the situation calls for deployment, it will be done in a way that respects the scientific evidence and protects the interests of the poor and vulnerable. Do we really believe that? The irony is that if we did believe in such a world, there would be no need for research into geoengineering. ■

Clive Hamilton is professor of public ethics at Charles Sturt University in Canberra and author of *Earthmasters: The Dawn of the Age of Climate Engineering* (Yale University Press, 2013).
e-mail@clivehamilton.com

**A SLEW OF PRIVATE
PATENTS
HAS ALREADY BEEN
TAKEN OUT
OVER METHODS TO
ENGINEER
THE WORLD'S CLIMATE.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/d59yjk

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

STEM CELLS

Telomeres help cells to commit

Stem cells have trouble assuming a specialized identity if their telomeres, the protective caps on the ends of their chromosomes, are short.

Pluripotent stem cells can become any cell type in the body, and are known to require long telomeres to grow normally. A group led by Lea Harrington, formerly at the University of Edinburgh, UK, reports that telomeres are also important for stem cells to differentiate into other cell types. They found that mouse embryonic stem cells with experimentally shortened telomeres did not differentiate stably, and that the pluripotency gene *Nanog*, which is normally silenced during differentiation, was abnormally active in these cells. In addition, the chemical modifications to DNA that normally stabilize differentiation were disrupted throughout the genome.

Telomere length might also have a role in differentiation of cancer stem cells in the body, the authors suggest.

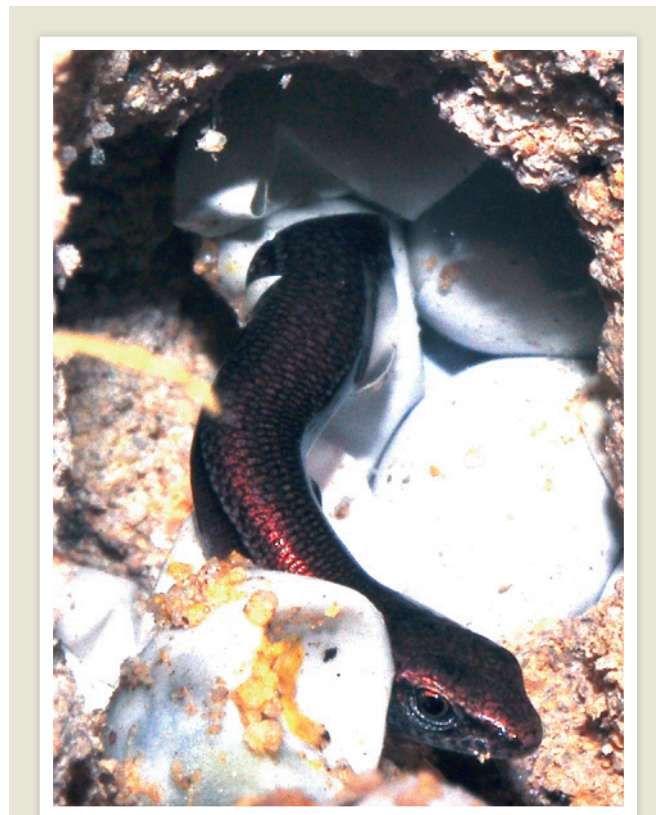
Cell Stem Cell 12, 479–486 (2013)

PSYCHOLOGY

Here's looking at you

When humans cannot tell where gaze is focused, they assume that people are looking at them.

Isabelle Mareschal, now at Queen Mary University of London, and her team asked volunteers to determine where the gazes of computer-generated faces were directed. When the true direction of the gaze was obscured by pixelating the eyes, all six observers



ANIMAL BEHAVIOUR

Lizards hatch early to flee

On sensing predators, developing delicate skinks (*Lampropholis delicata*, pictured) burst from their eggs and, in one fluid motion, sprint about 40 centimetres and dive for cover.

Sean Doody and Phillip Paull at Monash University in Clayton, Australia, poked, pinched and dropped skink eggs to simulate predator attacks. The duo found that threatened embryos escape their eggs several days earlier than unperturbed siblings. However, hatching early comes with a cost. Compared with spontaneously hatching skinks, hatchlings under threat leave more yolk behind and their bodies are about 4% shorter.

Although similar behaviour has been reported in amphibians, fish and invertebrates, this is the first report of reptiles hatching early in response to perceived predation.

Copeia 2013, 159–164 (2013)

believed the gaze was directed towards them — even if the eyes or faces were rotated to the left or right.

Although it is not clear whether this gaze bias is learned or innate, it could prove

useful in anticipating social interactions. Conditions such as autism have been linked to abnormal gaze behaviours, the authors note.

Curr. Biol. <http://dx.doi.org/10.1016/j.cub.2013.03.030> (2013)

SOLID-STATE PHYSICS

Diamonds tick like atomic clocks

Flaws in diamond crystals could make precise timekeeping more convenient.

Today's most accurate timekeeping standards are kept by clocks that contain hard-to-manufacture atomic gases. These clocks are usually placed only in specialist laboratories or on satellites, where their signals are used for applications such as communication and navigation. Dirk Englund, now at Massachusetts Institute of Technology in Cambridge, and his colleagues propose a scheme for keeping time using a diamond containing a nitrogen impurity. This defect has an oscillating electronic spin state that could be detected from the light it emits when excited by a laser, and therefore could provide a timekeeping signal. A device that relies on diamond chips would be more portable than atomic clocks, as well as easier to integrate into solid-state manufacturing.

Phys. Rev. A 87, 032118 (2013)

GEOLOGY

Quake linked to drilling

Waste water from oil and gas drilling that was pumped underground may have set off a magnitude-5.7 earthquake in the central United States.

In November 2011, a series of quakes near Prague, Oklahoma, could be felt across 17 states. A team led by Katie Keranan at the University of Oklahoma in Norman analysed the aftershocks to see how faults ruptured. Of three segments to break in the fault network, the first was located

NADAV PEZARO

within 200 metres of a shaft that had been used to inject fluid underground for 18 years. The authors conclude that changes in ground stress as the fluid built up probably caused the faults to fail.

Wastewater injections are known to have triggered earthquakes before, but this is the largest that has been tied to the practice so far, the authors say. They suggest that even a relatively small volume of fluid can lead to significant quakes many years down the line.

Geology <http://dx.doi.org/10.1130/G34045.1> (2013)

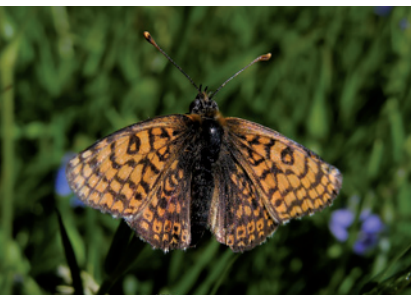
METABOLISM

Butterflies that live fast, die old

Butterflies with speedier metabolisms may live longer.

Kristjan Niitepõld at Stanford University in California and Ilkka Hanski at the University of Helsinki measured the energy expenditure of Glanville fritillary butterflies (*Melitaea cinxia*, pictured) by placing them in chambers and recording their carbon dioxide output. Every few days, the butterflies were forced to fly in the chambers for 10-minute stretches. Butterflies with the highest rates of CO₂ production lived the longest. Field experiments in which butterflies were released and recaptured showed the same pattern. Peak metabolic rate during flight could explain up to 33% of lifespan variation.

Such results are at odds with the generalization that faster metabolisms accelerate ageing and contribute to early death. **J. Exp. Biol.** 216, 1388–1397 (2013)



PAVE VASANEN

CANCER BIOLOGY

Targeting cancer metabolism

Compounds that inhibit mutated metabolic enzymes may shift cancer cells into a more benign state.

Mutations in two metabolic genes — *IDH1* and *IDH2* — drive the growth of some tumours. Katharine Yen at Agios Pharmaceuticals in Cambridge, Massachusetts, Ingo Mellingerhoff at the Memorial Sloan-Kettering Cancer Center in New York and their colleagues found a compound that selectively inhibits a mutated form of the IDH1 protein. In cells that carry these mutations, the inhibitor blocked formation of a cancer-associated metabolite. The compound also delayed cell growth and prompted differentiation.

Another team led by Yen designed a compound that inhibits mutated IDH2 protein. In human leukaemia cells that bear this mutation, the compound reduced accumulation of the metabolite and prompted the cells to differentiate and mature, a transition thought to prevent cancerous growth. **Science** <http://dx.doi.org/10.1126/science.1234769>; <http://dx.doi.org/10.1126/science.1236062> (2013)

CELL BIOLOGY

Prions prompt multicellularity

Prions, pathogens that trigger catastrophic chain reactions of misfolded proteins in various diseases, can help yeast to survive in harsh conditions.

Yeast assume complex multicellular structures when resources are scarce, such as long-branching stalks when cells are starved of nitrogen, or ridges when fermentable carbon is in short supply. Researchers led by Randal Halfmann at the University of Texas Southwestern Medical Center in Dallas report that

COMMUNITY CHOICE

The most viewed papers in science

SINGLE-MOLECULE DYNAMICS

Cell motors wobble to binding sites

HIGHLY READ
on www.cell.com
19 Mar–2 April

Improved high-speed imaging of single molecules shows protein motors use random motions to clamber towards their targets.

When muscles contract or cells divide, the necessary movements depend on foot-like projections of the protein myosin 'stepping' their way to binding sites on strands of the protein actin. Researchers led by Yale E. Goldman at the University of Pennsylvania in Philadelphia attached a dye to a myosin subunit, then used superfast switching of polarized light from multiple directions to take microsecond-scale snapshots that revealed the subunit's orientation and rotations. After detaching from one binding site, myosin takes a forward step that propels it only about two-thirds of the distance towards the next site. Wild gyrations get it the rest of the way. The group is now using the same technique to reveal the dynamics of molecular motors involved in processes such as protein synthesis.

Biophys. J. 104, 1263–1273 (2013)

such adaptive formations occur when a protein called Mot3 forms prions, drastically changing its shape and causing other Mot3 proteins to convert to this state. The researchers show that a non-fermentable carbon source such as ethanol induces Mot3 to form prions, and a lack of oxygen eliminates them. Thus, prions can mediate how yeast cells act cooperatively in response to their environment.

Cell 153, 153–165 (2013)



TOM BEAN/GETTY

HYDROLOGY

More rain for the Central Plains

Climate change has altered the water cycle in the Upper Mississippi River Basin, producing rainier summers that send more runoff into the river (pictured).

Chris Frans at the University of Washington in Seattle and his team studied data from 1918 to 2007 and report that increasing precipitation during July and August, especially in the northwestern part of the region, raised stream flows in the Upper Mississippi.

Hydrological modelling

showed that although an expansion in the amount of land cleared for agriculture did increase runoff in some parts of the basin, these changes were not enough to explain the overall trend.

The findings suggest that climate change could hamper efforts to reduce the flow of nitrogen-rich runoff from the river into the Gulf of Mexico, where the nutrients create low-oxygen 'dead zones'.

Geophys. Res. Lett. <http://dx.doi.org/10.1002/grl.50262> (2013).

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

Silencing probe

Canada's information commissioner, Suzanne Legault, is to investigate claims that scientists are being silenced in seven federal agencies, including the departments of the environment, fisheries and oceans, and natural resources, and the National Research Council of Canada. Democracy Watch, a non-profit group based in Ottawa, announced the probe on 1 April. It follows a complaint jointly filed by the group and the University of Victoria's Environmental Law Centre Clinic in February, arguing that government policies prevent the media and the Canadian public from speaking to government scientists for news stories. See go.nature.com/jtvb9v for more.

Gender gap closing

The proportion of female scientists in the European Union is slowly increasing, statistics published on 5 April show. Women filled some 11% of top research positions in 2010, slightly up from 8% in 2002, finds the European Commission report *She Figures 2012*. The proportion of female PhD graduates rose by 2% to 35% in the same period. The report calls for further initiatives to close the gender-equality gap.

Morning-after pill

A senior district judge in New York has ordered the US Food and Drug Administration (FDA) to make a controversial emergency contraceptive pill available to girls under 17 without a doctor's prescription. In 2011, the administration of President Barack Obama overruled FDA plans to provide under-17s with access to the Plan B One-Step contraceptive pill

(levonorgestrel) without a prescription. In a 59-page ruling made on 4 April, Judge Edward Korman said that the government's decision was "arbitrary, capricious, and unreasonable". Allison Price, a spokeswoman for the US Department of Justice, said that the department "is reviewing the appellate options and expects to act promptly." See page 138 for more.

EVENTS

Preprint server

The open-access journal *PeerJ* has launched a preprint server (*PeerJ PrePrints*) for biological and medical sciences, onto which authors can upload draft or final

versions of articles. Authors must be paid-up members of *PeerJ* to submit more than one paper per year. The service is intended to stimulate a culture of preprint sharing among biomedical researchers who, unlike physicists and mathematicians, do not widely use the most well-known preprint server, arXiv.

H7N9 avian flu

The human toll of H7N9 avian influenza had grown to 24 cases and eight deaths as *Nature* went to press. The events were strung along China's eastern coast and centred on Shanghai, where five of the deaths have occurred. Health authorities have examined hundreds of

atmospheric composition. The project will measure biomass to assess terrestrial carbon stocks and fluxes including those of sub-boreal forest species such as *Fagus sylvatica* (pictured), which have an important role in the terrestrial carbon cycle. ESA is expected to announce its final decision in May; once approved, BIOMASS could launch around the end of the decade. See go.nature.com/rk1nl6 for more.



GERHARD SCHULZ/GETTY IMAGES

ESA puts focus on biomass

A €420-million (US\$547-million) radar project designed to measure global forest biomass in the finest detail yet is the prime contender for Europe's next Earth-observation mission. Following a meeting last month (see *Nature* 495, 15; 2013), science advisers to the European Space Agency (ESA) recommended funding the BIOMASS project over satellite initiatives that would have tracked global snow cover and

contacts of infected people, but not yet found any evidence that the virus is being passed from person to person. See page 145 for more on efforts to understand the virus.

FACILITIES

Lakes funding ends

Funding for research projects at the Experimental Lakes Area in Ontario, Canada, dried up this week, and scientists are no longer allowed to enter the area. Having announced in May 2012 that the freshwater research station would close, the government is now in discussions with the International Institute for Sustainable Development, a think tank headquartered

MARCIA MCNUTT

in Winnipeg, Manitoba, over whether the institute will take over the facility. "It's up in the air as to what's going to happen," says Chris Metcalfe, an environmental toxicologist at Trent University in Peterborough, Canada, who was studying the ecological effects of silver nanoparticles at the lakes. See go.nature.com/s87rok for more.

PEOPLE

Campaigning on

Climatologist James Hansen stepped down from his post at NASA on 2 April to spend more time campaigning to reduce greenhouse-gas emissions. Hansen has directed the agency's Goddard Institute for Space Studies in New York since 1981. He was one of the first scientists to express concern that a modest rise in global temperature, of as little as 1°C above recent levels, could harm the planet. He also focused much of his effort into lobbying for stronger government action against climate change. See go.nature.com/5bvjq6 for more.

New Science editor

Marcia McNutt, former director of the US Geological Survey (USGS), will succeed Bruce Alberts as editor-in-chief of *Science*, the American Association for the Advancement of Science



announced on 2 April. McNutt (**pictured**) resigned from the USGS in February and will start at *Science* in June. A geophysicist by training, McNutt has also been chief executive of the Monterey Bay Aquarium Research Institute in Moss Landing, California. See go.nature.com/eyn4uu for more.

RESEARCH

Antimatter origins

The first results from the International Space Station's cosmic-ray detector, the Alpha Magnetic Spectrometer, were announced on 3 April. They confirm the presence of a mysterious excess of positrons, seen in our Galaxy by two earlier experiments, but fail to pin down its origin. The results hint that the antimatter may be produced from collisions of dark matter — which makes up around 80% of the matter

in the Universe — but could also originate from spinning superdense stars known as pulsars. See go.nature.com/sqo1s9 for more.

Tamiflu data open

Pharmaceutical giant Roche in Basel, Switzerland, will hand over data on its influenza drug Tamiflu to the Cochrane Collaboration, which conducts systematic reviews of health-care research and is based in Oxford, UK. The *British Medical Journal* (*BMJ*) announced the move on 4 April. The move follows an earlier attempt by Roche to allow wider access to its data, which the Cochrane Collaboration and the *BMJ* had said was inadequate.

Explorers chosen

NASA on 5 April chose the two projects that it will launch in 2017 under its Astrophysics Explorer Program — the agency's oldest continuously running programme. The Transiting Exoplanet Survey Satellite will search for planets passing in front of nearby bright stars. The Neutron Star Interior Composition Explorer is a cosmic X-ray detector that will sit on the International Space Station. NASA selected the winning projects from four proposals submitted in September 2012, concluding that they offered the "best scientific value and most

COMING UP

13–16 APRIL

More discussion about the Higgs boson features at the spring meeting of the American Physical Society in Denver, Colorado.

go.nature.com/evsjj6

15 APRIL

The US Supreme Court hears arguments over whether there should be limits on gene patents. See page 150 for more.

feasible development plans". See go.nature.com/nvbmy6 for more.

FUNDING

Action on the brain

US President Barack Obama on 2 April launched an ambitious project to probe the inner workings of the functioning human brain, calling it "the next great American project". Obama asked Congress to provide about US\$100 million in 2014 for the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative. The project is expected to cost billions of dollars, take at least 15 years to complete and require the development of new technologies. See go.nature.com/jcyzly for more.

US research rally

Thousands of researchers, patients and patient advocates gathered in Washington DC on 8 April to call on lawmakers to protect funding for medical research. The rally was in protest against upcoming cuts that will carve 5% from the budgets of US agencies, including the National Institutes of Health. Organizers estimate that 10,000 people attended the event. See go.nature.com/lcs7v6 for more.

NATURE.COM

For daily news updates see:

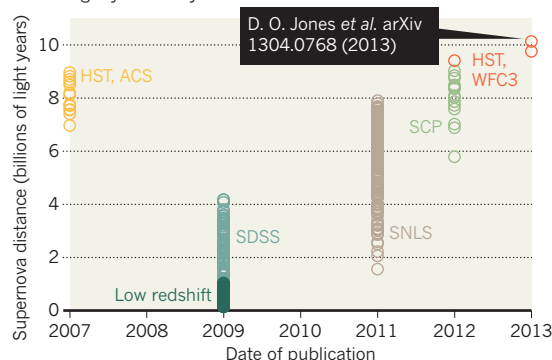
www.nature.com/news

TREND WATCH

The Hubble Space Telescope has spotted the farthest supernova yet seen of type Ia, which are used as 'standard candles' to measure cosmic distance. The 10-billion-year-old supernova UDS10Wil was spotted in 2010 by Hubble's near-infrared Wide Field Camera 3; it took three years for David Jones at Johns Hopkins University in Baltimore, Maryland and his colleagues to complete their analysis of the find, which is 350 million years older than the previous record holder.

RECORD FOR 'STANDARD CANDLE' SUPERNOVAE

Astronomers last week announced a type Ia supernova more than 10 billion light years away.



SCP: Supernova Cosmology Project; SNLS: SuperNova Legacy Survey; SDSS: Sloan Digital Sky Survey; low-redshift: assorted low-redshift discoveries; HST, ACS: Hubble Space Telescope with Advanced Camera for Surveys; HST, WFC3: Hubble Space Telescope with Wide Field Camera 3.

SOURCE: DAVID JONES, JOHNS HOPKINS UNIV.

NEWS IN FOCUS

CLIMATE Extreme weather events have adverse effect on carbon cycles **p.147**

MEDICINE Vets and pets embrace unproven stem-cell therapy **p.148**

NEUROSCIENCE Transparent technique illuminates whole brains **p.151**



GENETICS The hottest gene in translational medicine **p.152**

CHINA FOTOPRESS VIA GETTY



Chinese health workers swab a chicken to test for the novel H7N9 bird flu virus at a poultry market in Changsha, Hunan province.

INFECTIOUS DISEASE

Urgent search for flu source

Researchers suspect H7N9 virus is in bird markets as human cases rise rapidly.

BY DECLAN BUTLER

Virologists know its name: H7N9. What they don't yet know is whether this novel avian influenza virus — first reported in humans in China less than two weeks ago — will rapidly fizzle out, become established in animal hosts to fuel future human outbreaks, or morph into a virus that can spread easily between people and spark a deadly pandemic.

In a frantic effort to find answers, scientists are bearing down on H7N9 on multiple fronts. They are testing wild birds and thousands of domestic fowl; analysing the viruses they find;

and trying to trace people who have been exposed to infected patients. Chinese health authorities say that they have 400 laboratories looking for genetic changes in the virus.

“We are going to be sitting with bated breath over the next month to find out what happens,” says Michael Osterholm, who heads the University of Minnesota's Center for Infectious Disease Research and Policy in Minneapolis. As *Nature* went to press, 24 human cases, including 8 deaths, had been reported in 11 cities, some a few hundred kilometres apart, in eastern China (see map). So many cases in such a short time over such a wide area — up from three cases in two cities a week ago — is “a very

concerning situation”, says Osterholm.

Scientists urgently want to find out which sources are stoking the human infections that result in flu-like symptoms and, in most reported cases, severe pneumonia. So far, investigations of the cases remain largely inconclusive: some patients had contact with poultry or other animals just before falling ill, whereas others had not. Late last week, the H7N9 virus was found in chickens, pigeons and ducks in live bird markets in Shanghai and Hangzhou — making markets the leading suspected source. Authorities have since culled tens of thousands of birds and closed down markets in Shanghai, Nanjing and Hangzhou. ►



► The genetic sequences of the H7N9 viruses found in the birds are highly similar to those isolated from human patients, says Chao-Tan Guo, a virologist at the Zhejiang Academy of Medical Sciences in Hangzhou. Although the virus might have come from other sources, including mammals, the pattern of many human cases over a wide area in a short time could be explained by live markets alone, because birds from one or a few sources would be transported to multiple markets, says Malik Peiris, a flu virologist at the University of Hong Kong.

But the various bird species found to be infected may not be the original source, because much cross-infection can occur in live markets. Investigators must now trace which farms and wholesalers the birds came from, Peiris says, and test birds up through the supply chain.

Researchers know that H7 flu viruses mainly infect wild birds such as ducks, geese, waders and gulls, and that they occasionally jump into poultry flocks. Kwok-Yung Yuen, an infectious-disease expert at the University of Hong Kong, notes the proximity of the reported human cases to the Yangtze river delta, home to many wild birds, and to Chongming Island near Shanghai, a renowned site for watching migratory birds. “It’s likely wild ducks and

geese that are carrying it,” he suggests.

But this H7N9 virus has not yet been detected in wild birds in the area. “There is very little specific information on the source of this particular virus strain, its ecology or reservoir, and it is premature to be hypothesizing on the vectors,” says Taej Mundkur, who is flyways programme manager for conservation group Wetlands International in the Netherlands. He also co-convenes the Asia-Pacific Working Group on Migratory Waterbirds and Avian Influenza with the Food and Agriculture Organization of the United Nations (FAO).

HIDDEN RESERVOIR

Wherever the virus originated, a crucial question is whether it could become established in poultry, creating a reservoir that might lead to continued, sporadic human infections.

Health authorities in China are trying to learn to what extent that has happened already. Unlike its cousin H5N1 — which has killed millions of birds and several hundred people in Asia and elsewhere since 2003 — H7N9 does not cause serious bird disease, greatly complicating efforts to control it, says Vincent Martin, interim head of the FAO’s Emergency Prevention System for Transboundary Animal and Plant Pests and Diseases (EMPRES) in Rome. It would be next to

impossible to detect H7N9 through routine surveillance for sick poultry among China’s 6 billion domestic birds. “This means stopping animal-to-human transmission is impossible,” says Masato Tashiro, a virologist at the Influenza Virus Research Center in Tokyo, the World Health Organization’s influenza reference and research centre in Japan.

Each time the virus encounters new human hosts, it has fresh opportunities to mutate and to acquire the ability to spread between people. That does not seem to have occurred yet. But experts say that it will be crucial to identify and track new cases of suspicious severe pneumonia and their close contacts, and to isolate people if necessary. Researchers working on the molecular biology of the virus say that it seems to derive from a reassortment of genetic material from at least three known bird-flu groups (see *Nature* <http://doi.org/k4j>; 2013). A key component — the haemagglutinin (H) protein on the surface of the virus — already contains mutations known to shift its binding preference from bird cells to those of mammals. Scientists are watching for tell-tale changes that could signal a shift towards a form that is more transmissible between humans.

Because flu viruses evolve rapidly, comparing viral sequences from each of the human cases might reveal whether person-to-person transmission is occurring, says Andrew Rambaut, an expert in the evolution of human viral pathogens at the University of Edinburgh, UK. If many patients have very similar viral sequences, then that would imply human spread; if viral sequences are more diverse, it would imply that each person had separately picked up infections from birds. Only four sequences from four human cases are so far available, but virologists are sequencing more and posting them on the GISAID flu database.

If human-to-human transmission does start to occur, “further spread may be inevitable,” warns Tashiro. Humanity has never been widely exposed to H7 or N9 flu viruses, and so lacks resistance to these subtypes. If a pandemic were to occur, it would probably have a severe toll. But it is too early to predict how events will unfold; experts in emerging infectious disease are only just becoming acquainted with the latest villain in their roster. ■

Additional reporting by David Cyranoski.

MORE ONLINE

TOP STORY

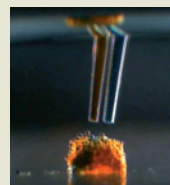


Hints of dark matter agree with earlier detections but remain inconclusive
go.nature.com/sqo1s9

MORE NEWS

- Quantum-entangled photons improve signal-to-noise ratio go.nature.com/kppvcr
- Bacteria turn red meat into artery-clogging compound go.nature.com/pwkawq
- Monarch butterflies migrate without a ‘map sense’ go.nature.com/yqlvvj

VIDEO



3D printer assembles artificial ‘living tissue’ from lipid-coated water droplets
go.nature.com/xwgnkn

CLIMATE

Wild weather can send greenhouse gases spiralling

Researchers get to grips with effects of heat, drought and storms on carbon release.

BY QUIRIN SCHIERMEIER

Climate change has a disconcerting tendency to amplify itself through feedback effects. Melting sea ice exposes dark water, allowing the ocean to soak up more heat. Arctic warming speeds the release of carbon dioxide from permafrost. And, as researchers discussed at a meeting last week in Seefeld, Austria, climate extremes — heatwaves, droughts and storms — can hamper plant growth, weakening a major buffer against the rise of CO₂ in the atmosphere.

“Heatwaves and droughts will very likely become more frequent in a warmer climate, and ecosystems will somehow respond,” says Philippe Ciais, a carbon-cycle researcher at the Laboratory of Climate and Environmental Sciences in Gif-sur-Yvette, France. “More storms will add an extra dimension to the problem.”

The meeting was organized by the CARBO-Extreme project, a €3.3-million (US\$4.5-million) collaboration of 27 groups from 12 countries, funded by the European Union. Attendees showed off an array of tools for uncovering how extreme events affect terrestrial carbon cycles, including numerical models, CO₂ flux measurements and field experiments. The challenge now, says Ciais, is to predict how the frequency of climate extremes will change, and to model the intricate physiological responses — some of which are poorly understood — of plants and ecosystems.

Land plants create a huge carbon ‘sink’ as they suck CO₂ out of the air to build leaves, wood and roots. The sink varies from year to year, but on average it soaks up one-quarter of the annual CO₂ emissions from the burning of fossil fuels. And events such as droughts, wildfires and storms are likely to “cause a pronounced decline” in the sink, says Markus Reichstein, a carbon-cycle scientist at the Max Planck Institute of Biogeochemistry in Jena, Germany, who coordinates CARBO-Extreme.

Climate anomalies have already had a detectable impact. Satellite observations and data from CO₂ measurement towers suggest that extreme events reduce plant productivity by an average of 4% in southern Europe and 1% in northern Europe, says Reichstein. That lowers annual carbon uptake by 150 million tonnes — equivalent to more than 15% of Europe’s annual man-made CO₂ emissions. The most extreme events can



A storm in southern France in 2009 left splintered trees and a weakened carbon sink.

turn forests and grasslands from carbon sinks to sources. In 2003 alone, a record-breaking heatwave in Europe led to the release of more CO₂ than is normally locked up over four years¹.

So far, scientists have detected no increase in extreme weather events. But they expect one. Reindert Haarsma, a climatologist at the Royal Netherlands Meteorological Institute in De Bilt, forecasts a surge in hurricane-strength storms such as 1999’s Lothar, which raged northeastward from the Bay of Biscay, slashing forest biomass by 16 million tonnes. By the end of this century, model studies suggest, storms similar to Lothar and another that caused huge damage in France in 2009 will become 25 times more common in Europe². The probability of major heatwaves in Europe is expected to increase up to tenfold by mid-century³.

Lack of water makes plants less capable of fending off pathogens and insects. After the 2003 heatwave, caterpillars devastated Mediterranean oak forests near Montpellier in France. Researchers have presumed that this triggered a large carbon release, but such responses are hard to predict. Severe droughts in 2005 and 2010 in the Amazon basin seem to have released much less CO₂ than expected, says Ciais. “In some ecosystems, small disturbances can have a large impact,” he adds. “In others, even significant anomalies seem to cause only little harm.”

CARBO-Extreme teams have conducted field experiments that simulated drought in different climates and vegetation types, from Atlantic pine forests to alpine meadows. Unpublished results show that in grasslands, drought markedly slowed photosynthesis, which stores carbon in leaves, roots and soil. It had a smaller effect on soil respiration, which releases carbon, so the net result was a decline in carbon uptake.

The experiments also showed that plants and soils keep a ‘memory’ of disturbances, says Michael Bahn, an ecologist at the University of Innsbruck in Austria who oversees a grassland experiment. He simulated a series of droughts, and found that later ones had a larger effect on net carbon release. Existing biosphere models do not capture such effects, which Bahn thinks might be due to changes in soil microbes.

Such omissions could lead to a large bias in the models. The world’s soils contain almost 100 gigatonnes of carbon — twice as much as the entire atmosphere. Just a 10% increase in soil-respiration rates, says Bahn, would release more CO₂ in a year than humans pump out. ■

1. Ciais, P. *et al.* *Nature* **437**, 529–533 (2005).
2. Haarsma, R. J. *et al.* *Geophys. Res. Lett.* <http://dx.doi.org/10.1002/grl.50360> (2013).
3. Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. & García-Herrera, R. *Science* **332**, 220–224 (2011).

JEAN-PIERRE MULLER/AFP/GETTY



A horse is given an injection of stem cells in a bid to promote healing in a tendon injury.

REGENERATIVE MEDICINE

Stem cells boom in vet clinics

Horses, dogs and even a tiger have received the unproven therapies. Now, drug regulators plan to weigh in.

BY DAVID CYRANOSKI

Patients seeking unproven stem-cell therapies in the United States often run up against government restrictions. But Vintage 'Vinty' Mark of Lovettsville, Virginia, had no difficulty getting such injections to treat an injured tendon in his leg. The leg improved dramatically, and Vinty went back to training — to be a racehorse.

New guidance from the US Food and Drug Administration (FDA) could, however, soon rein in veterinary uses of stem cells, a practice that has exploded in the United States over the past decade, even though most therapies are unproven. Many researchers and veterinarians say that the guidance, a draft of which the agency plans to issue by the end of the year, is overdue. But others worry that FDA interference could hamper research that could benefit animals — and their human companions.

In the absence of clear regulations, the industry has burgeoned. Vet-Stem, a company based in Poyay, California, has provided stem-cell treatments to more than 5,000 horses, 4,300 dogs and 120 cats since treating

its first patient in 2004. Kits provided by Medi-Vet America, based in Nicholasville, Kentucky, have been used to produce stem-cell injections for more than 10,000 horses since 2010. University veterinary departments, independently or through spin-off companies, have offered such services to thousands more animals. Veterinarians send patients' tissue samples to the centres to have cells extracted or, increasingly, turn to kits that allow them to extract the cells in-house.

Stem cells are most often used to treat horses, dogs and cats, but clinicians have also sought to use them to repair a lumbar fracture in a Bengal tiger and arthritis in pigs. Researchers have also found stem cells in the fat of bottlenose dolphins, raising hopes for treating the marine-mammal versions of liver disease and type 2 diabetes. "There's not a large vet practice that's not using them," says Wesley Sutter, a veterinarian at Lexington Equine Surgery and Sports Medicine in Kentucky. "Some claim [the treatment] cures everything."

➔ NATURE.COM
Read more about
stem-cell treatments
in humans:
go.nature.com/w3rqac

Many veterinarians offer unproven stem-cell therapies to satisfy demanding customers, says Dori Borjesson, who specializes in veterinary medicine at the University of California, Davis. "Clinicians are sucked into giving treatment" even when there's not research to back up uses, she says.

Like the treatments sought by humans, most of those used in animals involve mesenchymal stem cells (MSCs), which can mature into a wide variety of cell types, including bone and cartilage, and have been shown to have anti-inflammatory and other beneficial effects. MSCs are extracted from fat or bone marrow and can be cultured or prepared for injection in concentrated form.

The FDA's position on the use of MSCs in humans is clear. It says that the cells are drugs and therefore must be proved safe and effective before they can be used in treatment, except under certain conditions. No MSC treatments have been approved. But the FDA has different regulations for veterinary medicine, and these do not clearly address MSCs. The agency has not approved any veterinary stem-cell therapies, but neither has it cracked down on any. This is in stark contrast to its high-profile actions against purveyors of unproven human stem-cell treatments, such as Celltex Therapeutics of Sugar Land, Texas, which treated patients with MSCs until the FDA stepped in last September.

That doesn't mean that the agency is not concerned, says Lynne Boxer, a veterinary medical officer in the FDA's Office of New Animal Drug Evaluation in Rockville, Maryland. "As with any type of drug product, there are risks and benefits," she says. "With stem cells, there is the potential for disease transmission and tumour formation." She declines, however, to say whether current practices are against FDA rules, or to elaborate on what the new draft guidance is likely to contain.

The guidance void is irksome, says Karl Nobert, a lawyer at Squire Sanders in Washington DC, who has represented companies seeking direction from the FDA. His concern is not just professional: he is Vinty's former owner, and says that he saw "incredible improvement" as the horse healed, with normal tendon fibres rather than scar tissue at the injury site.

Research backs up the benefits of stem-cell treatments in some applications. A 2007 double-blind study in 21 dogs showed that MSCs improved chronic osteoarthritis¹. A 2010 report showed that injections of tissue rich in MSCs helped damaged leg bones to heal in 12 horses². And a 2012 study, much discussed by veterinarians, showed that MSCs from bone marrow helped racehorses with tendon injuries avoid re-injury³.

Researchers recognize that many studies of veterinary stem-cell treatments have a major weakness: they lack control groups or blinded evaluation, which are crucial to show whether the treatments truly make a difference. "It is

JESSICA CROSS/CORNELL UNIV.

exceptionally difficult to do blinded studies in our patients because the owners or trainers specifically seek an institution for stem-cell therapy," says Lisa Fortier, a veterinary surgeon at Cornell University in Ithaca, New York, and a co-author of the 2010 cartilage study.

Without controlled trials, comparisons between treated and untreated animals are difficult. Owners who have paid a few thousand dollars for a therapy are more likely to follow strict rehabilitation protocols than those who have not. They are also more likely to rush horses back into action, which would skew reported recovery time, a measure of a treatment's effectiveness. And placebo effects — on the owners — can be powerful. "The cat looks like hell to me but the owner says: 'She looks so great. I love stem cells,'" says Borjesson.

Many veterinarians think that studying stem-cell therapies in large animals, which are more physiologically similar to humans than the mice often used in pre-clinical research, could be a powerful model for medicine — if studies are done right. Borjesson is running studies (with "as much blinding as possible") of stem-cell treatments for dry eye, another eye condition called recurrent uveitis, and inflammatory bowel disease in clients' pet dogs. She is also collaborating with researchers in human regenerative medicine, so that her findings can inform their studies of experimental therapies. And David Frisbie, a veterinary surgeon at Colorado State University in Fort Collins, is working with doctors to apply lessons from animal stem-cell treatments to joint injuries in humans.

If the pending FDA guidance clearly designates animal stem cells as drugs, veterinarians and companies who want to deploy such treatments will first have to conduct clinical trials, which Frisbie estimates could cost at least US\$5 million each. Nobert, meanwhile, is optimistic that veterinary stem-cell companies and university researchers will find ways to "design creative regulatory approval strategies to streamline and expedite the review-and-approval process" — which might, in turn, influence the human stem-cell regulatory process.

Still, he worries that overly stringent FDA rules about, for example, how cells are characterized, could "make the ultimate commercial price unreasonable".

For Vinty, it's a moot point. He's had his stem cells. But even with his tendon fixed, he was too slow to race. He now gives riding lessons to children. ■

1. Black, L. L. et al. *Vet. Ther.* **8**, 272–284 (2007).
2. Fortier, L. A. et al. *J. Bone Joint. Surg. Am.* **18**, 1927–1937 (2010).
3. Godwin, E. E., Young, N. J., Dudhia, J., Beamish, I. C. & Smith, R. K. W. *Equine Vet. J.* **44**, 25–32 (2012).

POLICY

Budget forces tough look at biodefence

US supply stockpile for combating bioterror attacks and pandemics feels the strain of funding cuts.

BY ERIKA CHECK HAYDEN

Budget cuts mean painful choices for managers of the US Strategic National Stockpile, a medical repository initially designed to aid the country's response to terrorist attacks.

Congress created a national stockpile in 1998, and this has grown into an all-purpose resource that can deliver supplies to a disaster site within 12 hours. But the federal budget crisis has reduced the stockpile's funding by 18% over the past four years (see 'Running Low'), raising doubts about whether the United States can afford to be prepared for every pandemic or natural disaster, from bird flu to hurricanes.

At a 3 April meeting, advisers to the Department of Health and Human Services (DHHS) unanimously endorsed a report recommending that the stockpile rely on science to guide decisions about what threats are most likely, what supplies are needed to respond and whether local officials can actually use them. "The [stockpile] will be buying less. There's no doubt about it," said Ali Khan, who directs the public health and preparedness office at the US Centers for Disease Control and Prevention in Atlanta, Georgia, which oversees the stockpile.

Federal officials are struggling to weigh up competing priorities. State and local disaster-response agencies increasingly rely on the stockpile to compensate for cuts to their own budgets. But the national programme is also the only buyer for expensive, specialized treatments developed after the 2001 World

Trade Center and anthrax terrorist attacks. It is spending increasing amounts on therapies that are unlikely ever to be used.

"There has clearly been mission creep," said Steve Krug, director of emergency medicine at the Ann & Robert H. Lurie Children's Hospital of Chicago in Illinois at the 3 April meeting.

The DHHS has already sharply reduced orders for some items, including a new botulism antitoxin that costs US\$1,250 per dose. Such decisions alarm companies making biodefence products, says Maureen Hardwick, executive secretary of the Alliance for Biosecurity in Washington DC, which represents many of those firms. "If they're researching and developing things that will not be procured, that sends a negative signal," Hardwick says.

But just replacing expired medications already in the stockpile will exceed the repository's projected budget by 20% next year. And the programme will eventually be asked to restock other high-priced anti-bioterrorism drugs first purchased by a separate agency, the Biomedical Advanced Research and Development Authority (BARDA), which supports the development of medications considered priorities for biosecurity.

BARDA has purchased 2 million doses of an anti-smallpox drug for more than \$150 per dose, even though the stockpile already contains enough smallpox vaccine to immunize the entire US population.

The stockpile will also foot the bill for two biological products to neutralize the toxin produced by anthrax bacteria, first purchased by BARDA at costs of \$2,900–3,500 and \$8,100 per dose. Meanwhile, the country continues to invest in biodefence research.

Fewer purchases may mean that the United States will be less prepared for unlikely events, such as bioterrorist attacks, than officials would like. But other nations made this trade-off long ago. Only two other countries — Japan and Israel — are thought to have enough smallpox vaccine to vaccinate their entire populations. And no other country has anything like the anti-bioterrorism arsenal of the United States.

"There has to be a thoughtful analysis of what we're willing to do without, which is a very painful discussion in our society," Krug said at the meeting. ■

RUNNING LOW

Strong financial support for the Strategic National Stockpile has given way to harsh budget reality.



*2014 funding based on congressional authorization.

Gene patents in the dock

As US Supreme Court justices prepare to hear arguments in *Myriad Genetics* case, observers are debating the impact of the outcome on personal genomics.

BY HEIDI LEDFORD

When Daniel Weaver pitches Genformatic to potential investors, he feels obliged to note a future legal uncertainty. The two-year-old company, based in Austin, Texas, offers whole-genome sequencing and analysis to researchers and physicians, with plans to apply the technology to medical diagnostics. But Weaver fears that the company could become ensnared in a thicket of thousands of patents. “Who knows how much it would cost in legal fees just to sort through that?” he says.

Weaver and others in his line of business are looking to the US Supreme Court to prune that thicket. On 15 April, the court will hear arguments in a long-running lawsuit intended to answer one question: are human genes actually patentable? Yet the implications of the court's decision — expected by the end of June — may be narrower for business and medicine than many people hope and think. The case is limited to patents that cover the sequence of a gene, rather than methods used to analyse it (see ‘A plethora of patents’). “Symbolically, this case is a pretty big deal,” says Robert Cook-Deegan, a policy researcher at Duke University in Durham, North Carolina. “But the practical consequences of it are limited.”

The case, *Association for Molecular Pathology v. Myriad Genetics*, tackles the validity of patents owned by Myriad Genetics, a medical diagnostics company based in Salt Lake City, Utah, on isolated DNA that encompasses the human genes *BRCA1* and *BRCA2*. Certain forms of these genes increase the risk of breast, ovarian and other cancers. Myriad says that its patents are necessary to protect its investment in research. But physicians and patients charge that the intellectual-property restrictions have limited development of — and access to — medical tests based on the genes. In 2009, the American Civil Liberties Union and the Public Patent Foundation, both based in New York, sued Myriad. The case has been rumbling through the courts ever since.

To many in biotechnology, it has ramifications beyond specific genes. The case highlights concerns that a network of individual gene patents could threaten the future of personalized medicine and whole-genome sequencing by blocking companies and clinicians from reporting a patient's genetic risk factors for different diseases. “It's as if somebody had a patent on the

X-ray images of the pelvic region of a human being,” says Weaver. “You could administer the test, but you wouldn't be able to inform the patient about that region. It's crazy.”

By some estimates, the number of patents on human DNA is indeed extensive. In 2005, researchers reported that 20% of human genes had been patented¹. Two weeks ago, another team raised that estimate to at least 41% (ref. 2). But some dispute these numbers and their implications. Christopher Holman, a law

“Symbolically, this case is a pretty big deal. But the practical consequences of it are limited.”

professor at the University of Missouri-Kansas City, read through 533 of the 4,270 patents referenced in the 2005 study, and found that more than one-quarter were unlikely to limit genetic testing³. “The literature is full of this kind of problem,” he says.

His analysis was backed up by Nicholson Price, an academic fellow at Harvard Law School in Cambridge, Massachusetts, who found that few, if any, DNA patents would be infringed by companies or clinics sequencing whole genomes of individuals for medical insight⁴. Many, for example, apply only to the selective isolation of specific stretches of DNA, says Price, whereas whole-genome sequencing is an untargeted sweep of the entire genome.

Myriad's contested patents are part of a dying breed, says David Resnick, a patent attorney at the law firm Nixon Peabody in Boston, Massachusetts. They were filed in 1995, before much

of the human genome was sequenced and put into the public domain. Many other US gene patents issued before the human genome was sequenced are no longer enforced, because the companies that hold them have stopped paying maintenance fees. “This case is a conversation we should have had 20 years ago,” says Resnick. “It's moot now.”

Cook-Deegan thinks that whole-genome approaches may still be threatened if courts interpret patent claims broadly. Christopher Mason, a genomics researcher at Weill Cornell Medical College in New York, says that companies and clinics should not have to bear the risk of a court case. “If you're so sure those patents won't be a problem,” he says, “when I get sued, you'll pay my court fees.”

The irony is that even if Myriad's patents are ruled invalid, tests for mutations in the *BRCA1* and *BRCA2* genes may not become more widely available. Myriad's portfolio also includes patents on methods of analysing *BRCA* genes for links to cancer — and these are outside the scope of the current case. “If the Supreme Court says, ‘No, genes aren't patentable,’ what's going to change about that test?” asks Resnick. “Not one person is going to be able to get it that couldn't before.” ■

1. Jensen, K. & Murray, F. *Science* **310**, 239–240 (2005).
2. Rosenfeld, J. & Mason, C. E. *Genome Med.* **5**, 27 (2013).
3. Homan, C. M. *Soc. Sci. Res. Network* <http://dx.doi.org/10.2139/ssrn.1894715> (2011).
4. Price, W. N. *Cardozo Law Rev.* **33**, 1601–1631 (2012).

THE WIDER VIEW

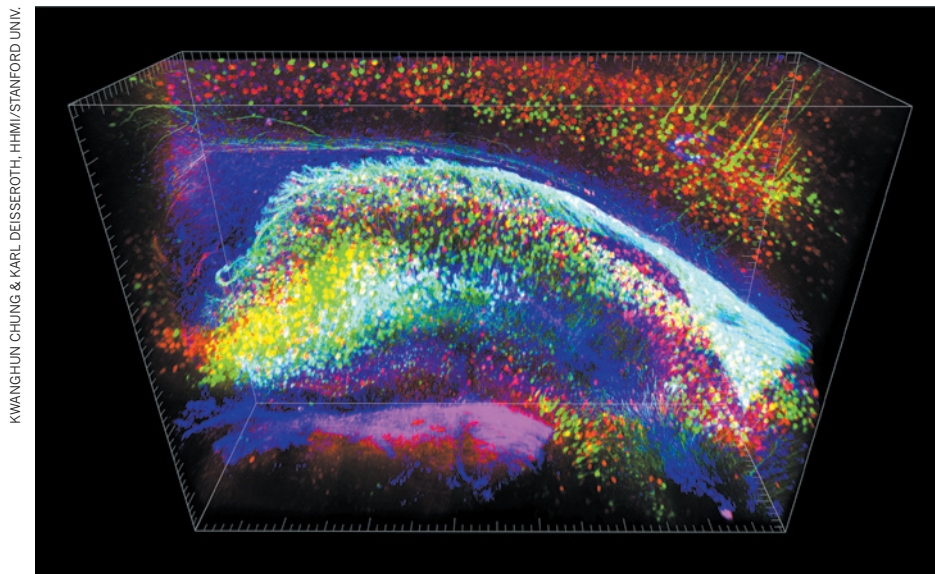
A plethora of patents

The phrase ‘gene patent’ is as ambiguous as it is emotionally charged. The US Supreme Court is set to evaluate whether genes that occur in nature can be claimed as innovations, but in so doing, it will focus on a narrow category of gene patent.

The patents in contention stake claims on isolated DNA sequences that make up *BRCA* genes. A federal court ruled in 2011 that isolating DNA changes it significantly from its natural state, rendering it fair game for a patent. (Legal scholars expect considerable time in arguments this month to be devoted

to analysing what ‘isolated’ means.)

But gene patents come in other flavours. Some are filed on engineered DNA sequences, others on gene variants linked to traits such as increased cancer risk. Yet others are filed on methods to determine whether a gene variant is present — blocking competitors from performing genetic tests. These are generally more vulnerable to court challenge than are patents that directly claim DNA sequences, and are easier to work around through new methodology. **H.L.**



Neurons in an intact mouse hippocampus visualized using CLARITY and fluorescent labelling.

NEUROSCIENCE

See-through brains clarify connections

Technique to make tissue transparent offers three-dimensional view of neural networks.

BY HELEN SHEN

A chemical treatment that turns whole organs transparent offers a big boost to the field of 'connectomics' — the push to map the brain's fiendishly complicated wiring. Scientists could use the technique to view large networks of neurons with unprecedented ease and accuracy. The technology also opens up new research avenues for old brains that were saved from patients and healthy donors.

"This is probably one of the most important advances for doing neuroanatomy in decades," says Thomas Insel, director of the US National Institute of Mental Health in Bethesda, Maryland, which funded part of the work. Existing technology allows scientists to see neurons and their connections in microscopic detail — but only across tiny slivers of tissue. Researchers must reconstruct three-dimensional data from images of these thin slices. Aligning hundreds or even thousands of these snapshots to map long-range projections of nerve cells is laborious and error-prone, rendering fine-grain analysis of whole brains practically impossible.

The new method instead allows researchers to see directly into optically transparent whole brains or thick blocks of brain tissue. Called CLARITY, it was devised by Karl Deisseroth

and his team at Stanford University in California. "You can get right down to the fine structure of the system while not losing the big picture," says Deisseroth, who adds that his group is in the process of rendering an entire human brain transparent.

The technique, published online in *Nature* on 10 April, turns the brain transparent using the detergent SDS, which strips away lipids that normally block the passage of light (K. Chung *et al. Nature* <http://dx.doi.org/10.1038/nature12107>; 2013). Other groups have tried to clarify brains in the past, but many lipid-extraction techniques dissolve proteins and thus make it harder to identify different types of neurons. Deisseroth's group solved this problem by first infusing the brain with acrylamide, which binds proteins, nucleic acids and other biomolecules. When the acrylamide is heated, it polymerizes and forms a tissue-wide mesh that secures the molecules. The resulting brain-hydrogel hybrid showed only 8% protein loss after lipid extraction, compared to 41% with existing methods.

Applying CLARITY to whole mouse brains, the researchers viewed fluorescently labelled neurons in areas ranging from outer layers of the cortex to deep structures such as the thalamus. They also traced individual nerve fibres

through 0.5-millimetre-thick slabs of formalin-preserved autopsied human brain — orders of magnitude thicker than slices currently imaged.

"The work is spectacular," says Van Wedeen, a neuroscientist at the Massachusetts General Hospital in Boston and a lead investigator on the US National Institutes of Health's Human Connectome Project (HCP), which aims to chart the brain's neuronal communication networks. The new technique, he says, could reveal important cellular details that would complement data on large-scale neuronal pathways that he and his colleagues are mapping in the HCP's 1,200 healthy participants using magnetic resonance imaging.

Francine Benes, director of the Harvard Brain Tissue Resource Center at McLean Hospital in Belmont, Massachusetts, says that more tests are needed to assess whether the lipid-clearing treatment alters or damages the fundamental structure of brain tissue. But she and others predict that CLARITY will pave the way for studies on healthy brain wiring, and on brain disorders and ageing.

Researchers could, for example, compare circuitry in banked tissue from people with neurological diseases and from controls whose brains were healthy. Such studies in living people are impossible, because most neuron-tracing methods require genetic engineering or injection of dye in living animals. Scientists might also revisit the many specimens in repositories that have been difficult to analyse because human brains are so large.

The hydrogel-tissue hybrid formed by CLARITY — stiffer and more chemically stable than untreated tissue — might also turn delicate and rare disease specimens into reusable resources, Deisseroth says. One could, in effect, create a library of brains that different researchers check out, study and then return. ■

CORRECTIONS

The News Feature 'The true cost of publishing' (*Nature* **495**, 426–429; 2013) wrongly described David Solomon as an economist; he is a psychologist. It also gave the wrong definition of STM — it should have said 'scientific, technical and medical'. Kay Dickersin's comment about jail in the News story 'Drug-company data vaults to be opened' (*Nature* **495**, 419–420; 2013) should have referred to the presentation of false results identified through transparency rather than failures to ensure transparency. The News story 'Crick's medal goes under the hammer' (*Nature* **496**, 13–14; 2013) should have said that Crick received his Nobel five, not six, decades ago. And the News Feature 'Fire in the hole!' (*Nature* **496**, 20–23; 2013) wrongly located Ted Jacobson at the University of Maryland, Baltimore, instead of in College Park.

A gene of rare effect

A mutation that gives people rock-bottom cholesterol levels has led geneticists to what could be the next blockbuster heart drug.

BY
STEPHEN S. HALL

When Sharlayne Tracy showed up at the clinical suite in the University of Texas (UT) Southwestern Medical Center in Dallas last January, the bandage wrapped around her left wrist was the only sign of anything medically amiss.

The bandage covered a minor injury from a cheerleading practice led by Tracy, a 40-year-old African American who is an aerobics instructor, a mother of two and a college student pursuing a degree in business. “I feel like I’m healthy as a horse,” she said.

Indeed, Tracy’s well-being has been inspiring to doctors, geneticists and now pharmaceutical companies precisely because she is so normal. Using every tool in the modern diagnostic arsenal — from brain scans and kidney sonograms to 24-hour blood-pressure monitors and cognitive tests — researchers at the Texas medical centre have diagnostically sliced and diced Tracy to make sure that the two highly unusual genetic mutations she has carried for her entire life have produced nothing more startling than an incredibly low level of cholesterol in her blood. At a time when the target for low-density lipoprotein (LDL) cholesterol, more commonly called ‘bad cholesterol’, in Americans’ blood is less than 100 milligrams per decilitre (a level many people fail to achieve), Tracy’s level is just 14.

A compact woman with wide-eyed energy, Tracy (not her real name) is one of a handful of African Americans whose genetics have enabled scientists to uncover one of the most promising compounds for controlling cholesterol since the first statin drug was approved by the US Food and Drug Administration in 1987. Seven years ago, researchers Helen Hobbs and Jonathan Cohen at UT-Southwestern reported¹ that Tracy had inherited two mutations, one from her father and the other from her mother, in a gene called *PCSK9*, effectively eliminating a protein in the blood that has a fundamental role in controlling the levels of LDL cholesterol. African Americans with similar mutations have a nearly 90% reduced risk of heart disease. “She’s our girl, our main girl,” says Barbara Gilbert, a nurse who has drawn some 8,000 blood samples as part of Cohen and Hobbs’ project to find genes important to cholesterol metabolism.

Of all the intriguing DNA sequences spat out by the Human Genome Project and its ancillary studies, perhaps none is a more promising candidate to have a rapid, large-scale impact on human health than *PCSK9*. Elias Zerhouni, former director of the US National Institutes of Health (NIH) in Bethesda, Maryland, calls *PCSK9* an “iconic example” of translational medicine in the genomics era. Preliminary clinical trials have already shown that drugs that inhibit the *PCSK9* protein — used with or without statins — produce dramatic reductions in LDL cholesterol (more than 70% in some patients). Half-a-dozen pharmaceutical companies — all aiming for a share of the global market for cholesterol-reducing drugs that could reach US\$25 billion in the next five years according to some estimates — are racing to the market with drugs that mimic the effect of Tracy’s paired mutations.

Zerhouni, now an in-house champion of this class of drug as an executive at drug firm Sanofi, headquartered in Paris, calls the discovery and development of *PCSK9* a “beautiful story” in which researchers combined detailed physical information about patients with shrewd genetics to identify a medically important gene that has made “super-fast” progress to the clinic. “Once you have it, boy, everything just lines up,” he says. And although the end of the *PCSK9* story has yet to be written — the advanced clinical trials now under way could still be derailed by unexpected side effects — it holds a valuable lesson for genomic research. The key discovery about *PCSK9*’s medical potential was made by researchers working not only apart from the prevailing scientific strategy of genome research over the past decade, but with an almost entirely different approach.

As for Tracy, who lives in the southern part of Dallas County, the implications of her special genetic status have become clear. “I really didn’t understand at first,” she admits. “But now I’m watching ads on TV [for cholesterol-lowering drugs], and it’s like, ‘Wow, I don’t have *that* problem.’”

A HEART PROBLEM

Cardiovascular disease is — and will be for the foreseeable future, according to the World Health Organization — the leading cause of death in the world, and its development is intimately linked to elevated levels of cholesterol in the blood. Since their introduction, statin drugs have been widely used to lower cholesterol levels. But Jan Breslow, a physician and geneticist at Rockefeller University in New York, points out that up to 20% of patients cannot tolerate statins’ side effects, which include muscle pain and even forgetfulness. And in many others, the drugs simply don’t control cholesterol levels well enough.

The search for better treatments for heart disease gained fresh impetus after scientists published the draft sequence of the human genome in 2001. In an effort to identify the genetic basis of common ailments such as heart disease and diabetes, geneticists settled on a strategy based on the ‘common variant hypothesis’. The idea was that a handful of disease-related versions (or variants) of genes for each disease would be common enough — at a frequency of roughly 5% or so — to be detected by powerful analyses of the whole genome. Massive surveys known as genome-wide association studies compared the genomes of thousands of people with heart disease, for example, with those of healthy controls. By 2009, however, many scientists were lamenting the fact that although the strategy had identified many common variants, each made only a small contribution to the disease. The results for cardiovascular disease have been “pretty disappointing”, says Daniel Steinberg, a lipoprotein expert at the University of California, San Diego.

More than a decade earlier, in Texas, Hobbs and Cohen had taken the opposite tack. They had backgrounds in Mendelian, or single-gene, disorders, in which an extremely rare variant can have a big — often fatal — effect. They also knew that people with a particular Mendelian disorder didn’t share a single common mutation in the affected gene, but rather had a lot of different, rare mutations. They hypothesized that in complex disorders, many different rare variants were also likely to have a big effect, whereas common variants would have relatively minor effects (otherwise natural selection would have weeded them out). “Jonathan and I did not see any reason why it couldn’t be that rare variants cumulatively contribute to disease,” Hobbs says. To find these rare variants, the pair needed to compile detailed physiological profiles, or phenotypes, of a large general population. Cohen spoke of the need to “Mendelize” people — to compartmentalize them by physiological traits, such as extremely high or low cholesterol levels, and then look in the extreme groups for variations in candidate genes known to be related to the trait.

The pair make a scientific odd couple. Hobbs, who trained as an MD, is gregarious, voluble and driven. Cohen, a soft-spoken

geneticist from South Africa, has a laid-back, droll manner and a knack for quantitative thinking. In 1999, they set out to design a population-based study that focused on physical measurements related to heart disease. Organized with Ronald Victor, an expert on high blood pressure also at UT Southwestern, and funded by the Donald W. Reynolds Foundation in Las Vegas, Nevada, the Dallas

Heart Study assembled exquisitely detailed physiological profiles on a population of roughly 3,500 Dallas residents². Crucially, around half of the participants in the study were African Americans, because the researchers wanted to probe racial differences in heart disease and high blood pressure. The team measured blood pressure, body mass index, heart physiology and body-fat distribution, along with a battery of blood factors related to cholesterol metabolism — triglycerides, high-density lipoprotein (HDL) cholesterol and LDL cholesterol. In the samples of blood, of course, they also had DNA from each and every participant.

As soon as the database was completed in 2002, Hobbs and Cohen tested their rare-variant theory by looking at levels of HDL cholesterol. They identified the people with the highest (95th percentile) and lowest (5th percentile) levels, and then sequenced the DNA of three genes known to be key to metabolism of HDL cholesterol. What they found, both in Dallas and in an independent population of Canadians, was that the number of mutations was five times higher in the low HDL group than in the high group³. This made sense, Cohen says, because most human mutations interfere with the function of genes, which would lead to the low HDL numbers. Published in 2004, the results confirmed that rare, medically important mutations could be found in a population subdivided into extreme phenotypes.

Armed with their extensive database of cardiovascular traits, Hobbs and Cohen could now dive back into the Dallas Heart Study whenever they had a new hypothesis about

heart disease and, as Cohen put it, “interrogate the DNA”. It wasn’t long before they had an especially intriguing piece of DNA at which to look.

THE MISSING LINK

In February 2003, Nabil Seidah, a biochemist at the Clinical Research Institute of Montreal in Canada, and his colleagues reported the discovery of an enigmatic protein⁴. Seidah had been working on a class of enzymes known collectively as proprotein convertases, and the researchers had identified what looked like a new member of the family, called NARC-1: neural apoptosis-regulated convertase 1.

“We didn’t know what it was doing, of course,” Seidah says. But the group established that the gene coding the enzyme showed activity in the liver, kidney and intestines as well as in the developing brain. The team also knew that in humans the gene mapped to a precise genetic neighbourhood on the short arm of chromosome 1.

That last bit of geographical information pointed Seidah to a group led by Catherine Boileau at the Necker Hospital in Paris. Her team had been following families with a genetic form of extremely high levels of LDL cholesterol known as familial hypercholesterolaemia, which leads to severe coronary artery disease and, often, premature death. Group member Marianne Abifadel had spent five



“The *PCSK9* story is a terrific example of an up-and-coming pattern of translational research.”

➔ **NATURE.COM**
Hear Stephen S. Hall discuss *PCSK9* on Nature's Podcast: go.nature.com/k9tear



MISTY KEASLER/REDUX/EVINE

Single-minded: Helen Hobbs and Jonathan Cohen's approach to heart-disease genetics yielded a target for drugs that could compete with statins.

fruitless years searching a region on the short arm of chromosome 1 for a gene linked to the condition. When Seidah contacted Boileau and told her that he thought *NARC-1* might be the gene she was looking for, she told him, "You're crazy," Seidah recalls. Seidah bet her a bottle of champagne that he was correct; within two weeks, Boileau called back, saying: "I owe you three bottles."

In 2003, the Paris and Montreal groups reported that the French families with hypercholesterolaemia had one of two mutations in this newly discovered gene, and speculated that this might cause increased production of the enzyme⁵. Despite Seidah's protests, the journal editors gave both the gene and its protein product a new name that fit with standard nomenclature: proprotein convertase subtilisin/kexin type 9, or *PCSK9*. At around the same time, Kara Maxwell in Breslow's group at Rockefeller University⁶ and Jay Horton, a gastroenterologist at UT-Southwestern⁷ also independently identified the *PCSK9* gene in mice and revealed its role in a previously unknown pathway regulating cholesterol⁸.

The dramatic phenotype of the French families told Hobbs that "this is an important gene". She also realized that in genetics, mutations that knock out a function are much more common than ones that amplify function, as seemed to be the case with the French families. "So immediately I'm thinking, a loss-of-function mutation should manifest as a low LDL level," she says. "Let's go and see if that's true."

GOING TO EXTREMES

Hobbs and Cohen had no further to look than in the extreme margins of people in the Dallas Heart Study. In quick order, they identified the highest and lowest LDL readings in four groups: black women, black men, white women and white men. They then resequenced the *PCSK9* gene in the low-cholesterol groups, looking for mutations that changed the make-up of the protein.

They found seven African Americans with one of two distinct 'nonsense' mutations in *PCSK9* — mutations that essentially

aborted production of the protein. Then they went back and looked for the same mutations in the entire population. Just 2% of all black people in the Dallas study had either of the two *PCSK9* mutations — and those mutations were each associated with a 40% reduction of LDL cholesterol in the blood⁹. (The team later detected a 'missense mutation' in 3% of white people, which impaired but did not entirely block production of the protein.) The frequency of the mutations was so low, Hobbs says, that they would never have shown up in a search for common variants.

When Hobbs and Cohen published their findings in 2005, they suggested that *PCSK9* played a crucial part in regulating bad cholesterol, but said nothing about whether the mutations had any effect on heart disease. That evidence came later that year, when they teamed up with Eric Boerwinkle, a geneticist at the University of Texas Health Science Center in Houston, to look for *PCSK9* mutations in the Atherosclerosis Risk in Communities (ARIC) study, a large prospective study of heart disease that had been running since 1987. To experts such as Steinberg, the results¹⁰ — published in early 2006 — were "mind-blowing". African Americans in ARIC who had mutations in *PCSK9* had 28% less LDL cholesterol and an 88% lower risk of developing heart disease than people without the mutations. White people with the less severe mutation in the gene had a 15% reduction in LDL and a 47% reduced risk of heart disease.

How did the gene exert such profound effects on LDL cholesterol levels? As researchers went on to determine¹¹, the *PCSK9* protein normally circulates in the bloodstream and binds to the LDL receptor, a protein on the surface of liver cells that captures LDL cholesterol and removes it from the blood. After binding with the receptor, *PCSK9* escorts it into the interior of the cell, where it is eventually degraded. When there is a lot of *PCSK9* (as in the French families), there are fewer LDL receptors remaining to trap and remove bad cholesterol from the blood. When there is little or no *PCSK9* (as in the black people with mutations), there

are more free LDL receptors, which in turn remove more LDL cholesterol.

The UT-Southwestern group, meanwhile, went back into the community looking for family members who might carry additional *PCSK9* mutations. In September 2004, Gilbert, the nurse known as ‘the cholesterol lady’ in south Dallas because of her frequent visits, knocked on the door of Sharlayne Tracy’s mother, an original member of the Dallas Heart Study. Gilbert tested Tracy, as well as her sister, brother and father. “They tested all of us, and I was the lowest,” Tracy says. Zahid Ahmad, a doctor working with Hobbs at UT-Southwestern, was one of the first to look at Tracy’s lab results. “Dr Zahid was in awe,” Tracy recalled. “He said, ‘You’re not supposed to be so healthy!’”

It wasn’t just that her LDL cholesterol measured 14. As a person with two dysfunctional copies of the gene — including a new type of mutation — Tracy was effectively a human version of a knock-out mouse. The gene had been functionally erased from her genome, and *PCSK9* was undetectable in her blood without any obvious untoward effects. The genomics community might have been a little slow to understand the significance, Hobbs says, “but the pharmaceutical companies got it right away”.

THE NEXT STATIN?

This being biology, however, the road to the clinic was not completely smooth. The particular biology of *PCSK9* has so far thwarted efforts to find a small molecule that would interrupt its interaction with the LDL receptor and that could be packaged in a pill. But the fact that the molecule operates outside cells means that it is vulnerable to attack by monoclonal antibodies — one of the most successful (albeit most expensive) forms of biological medicine.

The results of early clinical trials have caused a stir. Regeneron Pharmaceuticals of Tarrytown, New York, collaborating with Sanofi, published phase II clinical-trial results¹² last October showing that patients with high LDL cholesterol levels who had injections every two weeks of an anti-*PCSK9* monoclonal antibody paired with a high-dose statin saw their LDL cholesterol levels fall by 73%; by comparison, patients taking high-dose statins alone had a decrease of just 17%. Last November, Regeneron and Sanofi began to recruit 18,000 patients for phase III trials that will test the ability of their therapy to cut cardiovascular events, including heart attacks and stroke. Amgen of Thousand Oaks, California, has also launched several phase III trials of its own monoclonal antibody after it reported similarly promising results¹³. Among other companies working on *PCSK9*-based therapies are Pfizer headquartered in New York, Roche based in Basel, Switzerland, and Alnylam Pharmaceuticals of Cambridge, Massachusetts. (Hobbs previously consulted for Regeneron and Pfizer, and now sits on the corporate board of Pfizer.)

Not everyone is convinced that a huge market awaits this class of cholesterol-lowering drugs. Tony Butler, a financial analyst at Barclays Capital in New York, acknowledges the “beautiful biology” of the *PCSK9* story, but wonders if the expense of monoclonal drugs — and a natural reluctance of both patients and doctors to use injectable medicines — will constrain potential sales. “I have no idea what the size of the market may be,” he says.

“Everything hinges on the phase III side effects,” says Steinberg. So far, the main side effects reported have been minor, such as reactions at the injection site, diarrhoea and headaches. But animal experiments have raised potential red flags: the Montreal lab reported in 2006 that knocking out the gene in zebrafish is lethal to embryos¹⁴. That is why the case of Tracy was “very, very helpful” to drug companies, says Hobbs. Although her twin mutations

have essentially deprived her of *PCSK9* throughout her life, doctors have found nothing abnormal about her.

That last point may revive a debate in the cardiology community: should drug therapy to lower cholesterol levels, including statins and the anti-*PCSK9* medicines, if they pan out, be started much earlier in patients than their 40s or 50s? That was the message Steinberg took from the people with *PCSK9* mutations in the ARIC study — once he got over his shock at the remarkable health effects. “My first reaction was, ‘This must be wrong. How could that be?’ And then it hit me — these people had low LDL from the day they were born, and that makes all the difference.” Steinberg argues that cardiologists “should get off our bums” and reach a consensus about beginning people on cholesterol-lowering therapy in their early thirties. But Breslow, a former president of the American Heart Association, cautions against being too aggressive too soon. “Let’s start out with the high-risk individuals and see how they do,” he says.

Not long after Hobbs and Cohen published their paper in 2006, they began to get invited to give keynote talks at major cardiology meetings. Soon after, the genetics community began to acknowledge the strength of their approach. In autumn 2007, then-NIH director Zerhouni organized a discussion at the annual meeting of the institutes’ directors to raise the profile of the rare-variant approach and contrast it with genome-wide studies. “Obviously, the two approaches are opposed to each other, and the question was, what was the relative value of each?” says Zerhouni. “I thought the *PCSK9* story was a terrific example of an up-and-coming pattern of translational research” — indeed, he adds, “a harbinger of things to come”.

Hobbs and Cohen might not have found their gene if they had not had a hunch about where to look, but improved sequencing technology and decreasing costs now allow genomicists to incorporate the rare variant approach and to mount large-scale sweeps in search of such variants. “Gene sequencing is getting cheap enough that if there’s another gene like *PCSK9* out there, you could probably find it genome-wide,” says Jonathan Pritchard, a population biologist at the University of Chicago, Illinois.

“What was amazing to us,” says Hobbs, “was that the genome project was spending all this time, energy, effort sequencing people, and they weren’t phenotyped, so there was no potential for discovery. We didn’t understand, and couldn’t understand, why everybody wasn’t doing what we were doing. Particularly when we started making discoveries.” ■

Stephen S. Hall is a science writer in New York who also teaches public communication to graduate students in science at New York University.

1. Zhao, Z. *et al.* *Am. J. Hum. Genet.* **79**, 514–523 (2006).
2. Victor, R. G. *et al.* *Am. J. Cardiol.* **93**, 1473–1480 (2004).
3. Cohen, J. C. *et al.* *Science* **305**, 869–872 (2004).
4. Seidah, N. G. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 928–933 (2003).
5. Abifadel, M. *et al.* *Nature Genet.* **34**, 154–156 (2003).
6. Maxwell, K. N., Soccio, R. E., Duncan, E. M., Sehaye, E. & Breslow, J. L. *J. Lipid Res.* **44**, 2109–2119 (2003).
7. Horton, J. D. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 12027–12032 (2003).
8. Maxwell, K. N. & Breslow, J. L. *Proc. Natl Acad. Sci. USA* **101**, 7100–7105 (2004).
9. Cohen, J. *et al.* *Nature Genet.* **37**, 161–165 (2005).
10. Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr & Hobbs, H. H. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
11. Horton, J. D., Cohen, J. C. & Hobbs, H. H. *J. Lipid Res.* **50**, S172–S177 (2009).
12. Roth, E. M., McKenney, J. M., Hanotin, C., Asset, G. & Stein, E. A. *N. Engl. J. Med.* **367**, 1891–1900 (2012).
13. Koren, M. J. *et al.* *Lancet* **380**, 1995–2006 (2012).
14. Poirier, S. *et al.* *J. Neurochem.* **98**, 838–850 (2006).



MICHAEL SOHN/AP

GERMANY'S ENERGY GAMBLE

An ambitious plan to slash greenhouse-gas emissions must clear some high technical and economic hurdles.

BY QUIRIN SCHIERMEIER

In an industrial warehouse on the outskirts of Stuttgart, the sooty past is shaking hands with Germany's green-energy future. In one corner sits a relic destined for a museum: a cast-iron engine as big as a bus, which was used until the early 1970s to compress coal gas for lighting, cooking and heating. Nearby, a gleaming network of stainless-steel tanks uses electricity to create methane out of water and carbon dioxide.

This power-to-gas (P2G) pilot plant is the largest of its kind in the world, and the research behind it could help to propel Germany to the front of the race for cleaner energy. Developers of the €3.5-million (US\$4.5-million) project say that the P2G technology is an ideal way to

cope with the unreliable nature of solar and wind power. During sunny or breezy days, excess electricity can be used to make methane, which can be stored and then burned to generate power when the winds fail or the days turn dark.

The quest to develop P2G storage is part of Germany's ambitious *Energiewende*, or energy transition, a long-term plan to clean up the country's energy systems. Enshrined in law, the scheme aims to slash CO₂ emissions by replacing fossil fuels with renewable sources of energy. Germany hopes to generate at least 35% of its electricity from green sources by 2020; by 2050, the share is expected to surpass 80% (see 'Green growth').

The *Energiewende* — the world's most extensive embrace of wind and solar power as well as other forms of renewable energy — enjoys the support of all Germany's political parties and most of the population. It will probably continue whatever the results of the national election in September. Other nations are watching keenly to see how the experiment proceeds, and whether they should follow the German lead. "Germany's *Energiewende* can mobilize a global energy revolution," says Harry Lehmann, executive chairman of the World Council for Renewable Energies, who is based in Dessau, Germany.

To reach its goal, Germany is currently investing more than €1.5 billion per year

“Don’t overturn the *Energiewende*,” urges a supporter of Germany’s plan to shift rapidly towards renewable energy.

in energy research. One of its chief aims is to improve and build more storage systems, such as the Stuttgart P2G plant.

Another is extending and strengthening the electricity grid to wire up remote wind turbines and countless small photovoltaic installations. The research programme also seeks to improve the efficiency of energy production from sunlight, wind and biomass, and to encourage people to reduce energy consumption.

Most experts in Germany agree that the technical hurdles are surmountable. “We don’t need technological leaps to accomplish the *Energiewende*,” says Eberhard Umbach, president of the Karlsruhe Institute of Technology, who oversees the €500-million energy-research activities of Germany’s national research centres.

But the economic challenges are daunting, with the total costs of the *Energiewende* estimated to top €1 trillion. Europe’s deep financial crisis looms large over a project of that scale, warns Roger Pielke Jr, an environmental-policy researcher at the University of Colorado Boulder. “The German public has so far shown great willingness to pay for the transformation, but there will be limits to that willingness, especially if the economic climate gets rougher.”

MONEY AND POWER

The *Energiewende* is already visible to anyone travelling the German countryside. Expensive solar panels cover more than one million houses, farms and warehouses, thanks to generous subsidies. Along the motorways, clumps of wind turbines sprout on ridges, particularly in the breezy northern regions. The country’s green-energy portfolio is growing at a ferocious rate; last year, renewables supplied more than one-quarter of Germany’s gross electricity needs.

The primary motivation for the *Energiewende* is to combat climate change. By 2020, Germany aims to cut its greenhouse-gas emissions by 40% below 1990 production levels, and it hopes to achieve a reduction of at least 80% by 2050. The government set those goals in 2005, but the targets became much more ambitious in the aftermath of the Fukushima nuclear accident in Japan in March 2011.

That crisis spurred German Chancellor Angela Merkel to speed up a planned move away from nuclear power, which provided 25% of the country’s primary power in 2010. She closed eight nuclear plants immediately and pledged to shut the remaining nine by 2022. Suddenly, Germany needed to pick up the pace of its move towards renewable energy.

For German consumers, the costs of that shift are apparent in their monthly electricity bills. The statements include a litany of ‘shared

costs’ that are split by all households to fund the *Energiewende* — and result in some of the highest electricity prices in Europe. (Heavy industries are currently exempt from paying the surcharge.)

The shared costs are a mechanism for promoting green forms of energy, which are more expensive to produce than electricity from coal and natural gas. Germany’s Renewable Energy

GERMANY’S ENERGIEWENDE CAN MOBILIZE A GLOBAL ENERGY REVOLUTION

Act (EEG), the legal force behind the *Energiewende*, allows owners of solar panels and wind turbines to sell their electricity to the grid at a fixed, elevated price. Renewable-power producers cashed in an estimated €20 billion last year for electricity that was actually worth a mere €3 billion on the wholesale electricity market. The difference came out of the pockets of consumers.

The EEG, first passed in 2000, has helped Germany to install many more wind and solar power systems than most other developed nations. But the subsidies are causing some grotesque distortions in the electricity market — to the point at which energy companies are sometimes forced to sell conventional power at a loss.

The effects of Germany’s policies are also reverberating through the European energy market. One of Europe’s biggest energy providers, E.ON based in Düsseldorf, announced in January that it plans to close several gas-fired power stations across Europe that were operating at a loss, even though they are far less polluting than coal-fired plants. The “unmanaged growth” of renewable-energy sources is forcing gas-fired plants to spend too much time idle, E.ON boss Johannes Teyssen told shareholders.

At the same time, Germany is subject to the vagaries of outside forces, such as the rapid

expansion of natural-gas production in the United States, which has curbed domestic demand for coal. Excess US coal is now heading to Europe,

helping to fuel a resurgence of coal use in the United Kingdom and Germany, among other countries. With Germany’s imports of low-cost coal rising, the country’s greenhouse-gas emissions increased by almost 2% in 2012 — bucking a long-term decline. Last August, federal minister of the environment Peter Altmaier made clear that coal-fired plants will be needed “for decades to come” to ensure energy supplies. Germany is currently building some 11 gigawatts of coal-fired plants and its existing capacity of around 55 GW will not shrink as quickly as the country had planned.

The coal resurgence makes it unlikely that Germany will meet its 2020 emissions targets, says Pielke Jr. “You must accept the logic,” he says. “If you opt out of nuclear power, your near-term emissions will go up.”

But German officials say that even if emissions rise temporarily, that trend will turn around. State-of-the-art coal-powered plants, such as a new 2.2-GW facility near Cologne that burns lignite, or low-grade coal, will replace several older, less efficient plants. The net effect will be to reduce CO₂ emissions, says Altmaier.

To sustain that trend, however, the costs of green energy need to come down. “Subsidies have helped to get the renewable thing started, but sooner or later renewable energy must become economically self-sustaining,” says Brigitte Knopf, head of German and European energy strategies at the Potsdam Institute for Climate Impact Research.

That is where Germany is aiming many of its energy-research efforts, which will receive in excess of €3.5 billion from the government between 2011 and 2014. About €200 million is going towards developing and improving storage technologies.

“Renewable energy is a wonderful thing — provided you are able to store it at a large scale and distribute it efficiently,” says Frithjof Staif, managing director of the Center for Solar Energy and Hydrogen Research (ZSW) in Stuttgart, which operates the P2G plant. By making renewable energy more manageable and marketable, advanced storage technologies can ultimately help to reduce the cost of wind and solar power, he says.

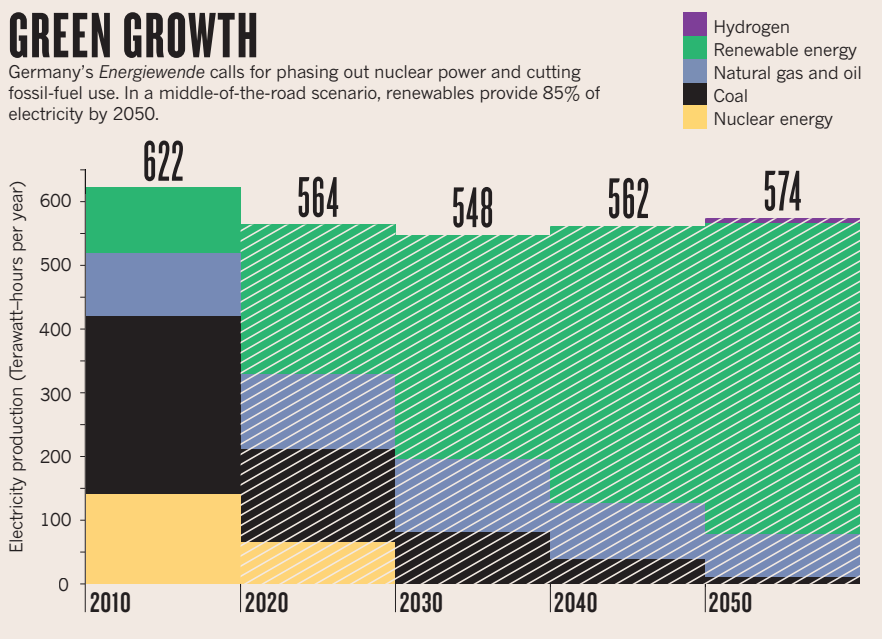
Germany currently stores excess electricity by using pumps to push water uphill into reservoirs. When electricity demand goes up, the water is released through turbines to generate hydroelectric power. But there are limits to expanding this type of storage. With 30 pump facilities already in operation, Germany has few suitable sites left for building more; the country plans to increase pump-storage capacity by just 20% by 2020.

P2G, however, could provide a vast amount of new storage capacity and Germany is leading the way. The plant in Stuttgart has 250 kilowatts of electrolysis stacks, which use electricity from renewables to produce hydrogen from water. To make methane, the hydrogen is reacted with

➔ **NATURE.COM**
For an interactive
guide to global
energy use, see:
go.nature.com/3em37o

GREEN GROWTH

Germany's *Energiewende* calls for phasing out nuclear power and cutting fossil-fuel use. In a middle-of-the-road scenario, renewables provide 85% of electricity by 2050.



CO₂ from decomposing sewage and agricultural waste at a nearby biogas plant. Other P2G plants could scrub CO₂ from the air.

But P2G is still an immature technology, with high upfront costs and an efficiency of only about 50% in converting electricity to methane. Synthetic methane plants have also struggled with the purity of their product. At the ZSW facility, the main goal is to routinely produce gas with low oxygen and hydrogen content.

Several other government-funded research groups are working to improve the membranes used in electrolysis reactions to produce the hydrogen for P2G plants. Advanced proton-exchange membranes, for example, can respond to electricity fluctuations within milliseconds, which could produce the very pure hydrogen needed for P2G and other clean-energy technologies.

Once P2G plants work out such issues, they could scale up, and synthetic gas could go directly into the network of pipelines and storage tanks currently used for natural gas. It could then be burned in power plants to produce electricity; by some estimates, this technology will be able to store up to 500,000 gigawatt-hours of electricity, enough to power Germany for more than six months.

The synthetic gas could be also used for heating or as a vehicle fuel. This prospect has piqued the interest of the car-maker Audi, which is currently building the first commercial P2G test facility in Werlte, northern Germany. The facility is expected to produce up to 4,000 cubic metres of synthetic methane per day — enough to fuel about 1,500 cars. Audi is also set to introduce a hybrid natural-gas-powered sedan later this year.

If Germans embrace this kind of vehicle, it could help to cut carbon emissions from the transport sector, which has lagged behind

others in implementing the *Energiewende*. The government intends to have one million battery-powered cars on German roads by 2020, but experts view that goal as unrealistic. The industry lacks the capacity to produce that many electric cars, and motor-loving Germans have not shown much desire for them. The new Audi and similar models may turn more heads, because consumers can switch to normal fuel if they are running low on natural gas and cannot find a specialized refuelling station. That option is not available to owners of battery-powered vehicles that run low on charge.

SUPPLY AND DEMAND

Although transportation has been slow to change, the electricity sector has the opposite problem. The rapid rise in wind and solar power has created a nightmare scenario for grid operators, who face power surges when the wind blows and the Sun shines, and shortages when they don't. In 2011, more than 200,000 blackouts exceeding three minutes were reported — and experts warn of a growing risk of major power failures.

For the *Energiewende* to succeed, the grid must be able to accommodate millions of extra small solar installations and wind turbines, as well as autonomous sub-grids such as those that connect offshore wind farms, which intermittently send floods of power into the onshore grid.

In January, the government put out a €150-million call for research proposals for improving the electricity network. The government also announced last year that it would install almost 4,000 kilometres of high- and low-voltage power lines, with a total transmission capacity of 10 GW. The €20-billion project would help to carry energy to the south of Germany from wind farms in the north.

And Germany may look even farther afield. Having distant power sources outside the nation would facilitate the *Energiewende* because a calm, dark day in Germany could be balanced by energy from other countries, says Robert Socolow, an environmental economist at Princeton University in New Jersey.

Some companies have floated plans to build large thermal solar plants in the Sahara Desert, which gets enough sunshine to meet most of Europe's electricity demand. But this scheme, the multibillion-euro DESERTEC initiative, lost momentum late last year, when two major industry partners — Siemens and Bosch — backed out (see *Nature* 491, 16–17; 2012). Energy analysts, moreover, doubt that Germany or any other European country would be willing to rely on substantial electricity imports from a politically unstable region.

Some question the need to ship power over long distances. "That the *Energiewende* depends on a huge expansion of the grid is a myth — and a very expensive myth, too," says Mathias Willenbacher, co-founder and managing director of juwi in Wörrstadt, a company that has successfully built small-scale renewable-energy projects. Willenbacher argues that it would be better to plough money into energy-storage options. "If you solve the storage issue, there is no need any more to transmit massive amounts of wind power from the North Sea to the Alps," he says.

However Germany sources its electricity, the *Energiewende* will not succeed unless the country can convince ordinary people to use energy more efficiently, says Umbach.

Researchers from the Karlsruhe Institute of Technology are currently running a field study to see whether they can alter the behaviours of 1,000 private citizens and commercial electricity customers in the city of Göppingen and the rural municipality of Freiamt. The trial, supported by the energy company EnBW in Karlsruhe, gives consumers smart-grid meters that provide information on their minute-to-minute energy use as well as the cost of power, which fluctuates throughout the day. The researchers aim to study whether people use the information to reduce their energy consumption.

How Germans will get by in the brave new energy world remains to be seen, says Umbach. "Will a computer tell them when to wash and cook?" he asks. "Or will they still want to fry their schnitzel when they feel like it?"

Surveying all the impediments to the *Energiewende*, Umbach is not sure that the transformation will come off as soon as planned. But he is convinced that Germany is the right country to try the great experiment. "If it fails it will be bad for Germany," he says. "But if it succeeds the whole world will profit." ■ [SEE EDITORIAL P.137](#)

Quirin Schiermeier is a reporter for *Nature* in Munich, Germany.

COMMENT

HISTORY In praise of Alfred Russel Wallace, visionary scientist, daring explorer **p.162**

BIOGEOGRAPHY David Quammen reappraises *The Malay Archipelago* **p.165**

MUSEUM San Francisco Exploratorium has a bright new home **p.167**

TEXTILES Knitted robots, mobile-charging shirts and wearable sensors **p.168**



PESKIMO/SYNERGY ART

A jump-start for electroceuticals

Kristoffer Famm and colleagues unveil a multidisciplinary initiative to develop medicines that use electrical impulses to modulate the body's neural circuits.

Imagine a day when electrical impulses are a mainstay of medical treatment. Your clinician will administer 'electroceuticals' that target individual nerve fibres or specific brain circuits to treat an array of conditions. These treatments will modulate the neural impulses controlling the body, repair lost function and restore health. They could, for example, coax insulin from cells to treat diabetes, regulate food intake to treat

obesity and correct balances in smooth-muscle tone to treat hypertension and pulmonary diseases.

All this is within reach if researchers from disparate disciplines in academia and industry work together. Here, we outline what needs to be done to bring about electroceuticals and unveil a public-private research initiative and an award that we hope will catalyse the field.

Electrical impulses — action potentials — are the language of the body's nervous system. Virtually all organs and functions are regulated through circuits of neurons communicating through such impulses¹. Two features make these circuits excellent targets for therapeutic intervention. First, they comprise discrete components — interconnected cells, fibre tracts and nerve bundles — allowing for pinpoint ►

► intervention. Second, they are controlled by patterns of action potentials, which can be altered for treatment.

Already, devices that harness electrical impulses are used to treat disease. Pacemakers and defibrillators save millions of lives each year; deep-brain stimulation dramatically improves the quality of life for people with Parkinson's disease and depression; sacral-nerve stimulation restores some bladder control in people with paraplegia, and vagus-nerve stimulation shows clinical benefits in diseases ranging from epilepsy to rheumatoid arthritis². But these devices do not target specific cells within circuits.

Neural tissue is compact: unrelated circuits often run close together through brain regions and in peripheral nerves. At present, electrical devices activate or inhibit cells in an area of tissue indiscriminately, muddying clinical effects. For example, electrodes that stimulate the vagus nerve enclose approximately 100,000 fibres, which innervate many different internal organs. Similarly, deep-brain stimulation for Parkinson's disease affects many cells other than those that control movement, leading to emotional and cognitive side effects. In natural urinary control, opposite signals in adjacent nerve fibres simultaneously contract the bladder and relax the urethral sphincter — an elegant process that is poorly mimicked by today's devices.

Neither do neurostimulation devices yet generate naturalistic patterns of action potentials. Typically, devices block or stimulate with simple waveforms, rather than modulate dynamically on the millisecond scale. Precise modulation is important: in mice, stimulation of cells in the neural circuit for hunger with a simple 20-Hertz waveform causes voracious eating within minutes, and ablation of these cells causes anorexia; but food intake can be more finely modulated by the number and frequency of action potentials in specific cells³. Similarly, single action potentials in small sets of cortical neurons have been shown to encode sensory input or perception in mice⁴. In other words, neural circuits act through sets of precise electrical impulses generated in specific sets of cells.

PATH TO PRECISION

We believe that it is now possible to create medicines that control action potentials in individual neurons and in functional groups of them.

Many of the stepping stones are already in place, thanks to recent advances in a variety of disciplines. For example, disease-specific neural circuits, such as the reflex that controls levels of inflammatory mediators⁵, are starting to be anatomically and functionally

traced. Tools, such as optogenetics, that enable cellular-level control have improved researchers' ability to analyse the signals in circuits, and they provide a mechanism by which future electroceuticals could elicit action potentials⁶. Efforts to control prosthetic limbs and generate brain-machine interfaces are giving rise to architectures for electrodes that can interact with individual neurons. Researchers are designing microchips that mimic brain processing to facilitate local and low-power computation⁷. The development of cochlear and retinal implants has led to advances in neural signal processing. Nanotechnology has delivered approaches for harvesting energy to power microdevices⁸. And neurosurgery can now be done through small holes in the skull and body with the use of needles and

"Researchers will need to embrace the tools of other fields, and even dream differently."

scopes, as in precision procedures to remove herniated disc material from the spine or open new fluid channels from blocked brain ventricles.

The first logical step towards electroceuticals is to better map the neural circuits associated with disease and treatment. This needs to happen on two levels. On the anatomical level, researchers need to map disease-associated nerves and brain areas and identify the best points for intervention. On the signalling level, the neural language at these intervention points must be decoded, so that researchers can develop a 'dictionary' of patterns associated with health and disease states — a project synergistic with international drives to map the human brain⁹. In circuits altered by disease, it will be important to establish how introduced electrical impulses affect the disease and which patterns yield the most effective therapeutic responses. Developing the technology to record from and stimulate a larger set of central and peripheral neurons will be crucial to this pursuit.

This type of research is analogous to the target-identification and validation steps at the core of modern molecular-drug discovery. The circuit maps that emerge will provide the design specifications for future treatment devices. Early prototypes might use microchip-controlled electrode arrays similar to those used today in interfaces for prosthetic limbs to modulate neural signals (see 'It's electric'). Second-generation micro- and nanoscale devices may instead leverage light, mechanical or magnetic energy to achieve such modulation in specific cells within targeted circuits.

How will all this come about? Disease biologists will need to work with neuroscientists to map circuits and with bioinformaticians

to identify the action-potential signatures of diseases. To develop treatment devices, bioengineers designing biocompatible interfaces will need to collaborate with electrical engineers to develop microchips for real-time signal processing; with nanotechnologists to create energy sources; and with neurosurgeons to ensure that these designs can be implanted and connected. Researchers will need to embrace the languages and tools of other fields, and perhaps even dream differently: much of the challenge lies in translating biological understanding into engineering specifications.

MULTIDISCIPLINARY JOURNEY

We think that initial progress will come from targeting circuits that have accessible and peripheral intervention points. For example, it has been shown that hypertension can be controlled through signals in carotid-sinus and renal nerves, and the production of certain inflammatory molecules in rheumatoid arthritis can be modified through the splenic nerve. A range of conditions — cardiovascular, metabolic, respiratory, inflammatory and autoimmune — are likely to have similarly accessible intervention points, given that they involve organs and functions that are under neural control.

We envision adaptive or 'closed-loop' electroceuticals that can record incoming action potentials and physiological parameters, analyse these data in real time and modulate neural signalling accordingly. This capability, together with that of spatially targeting a specific set of neurons, will underpin the selective therapeutic effect that we expect from electroceuticals. But these closed-loop therapies can be realized only if the required disciplines come together early on.

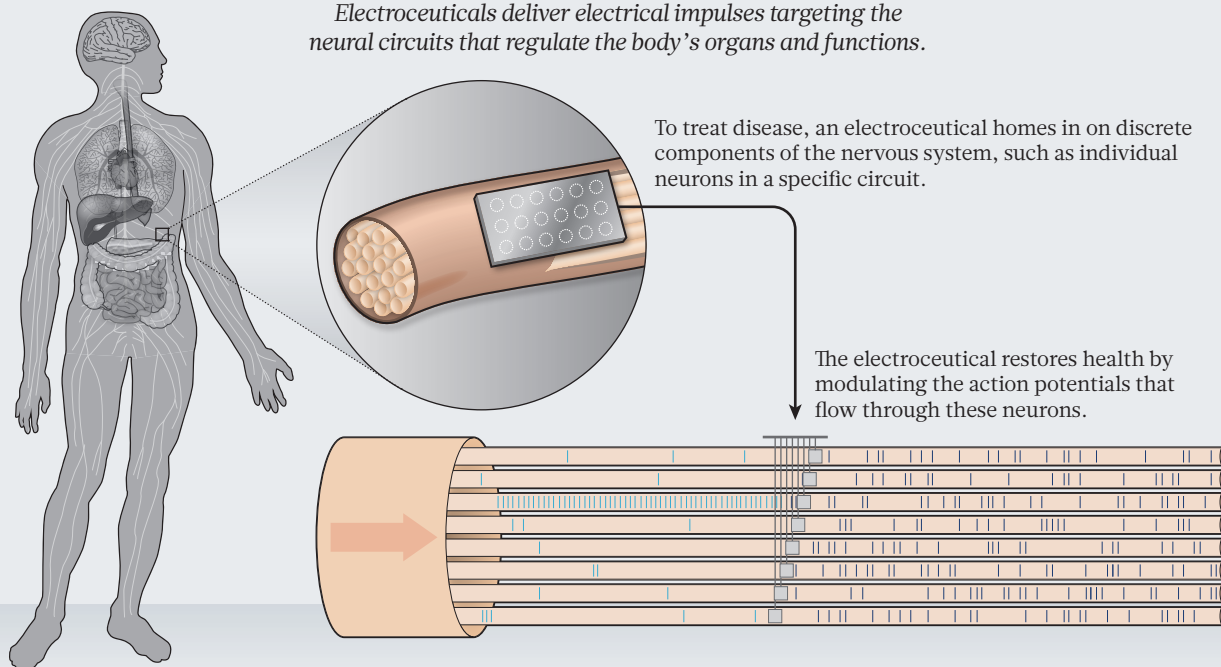
We also hope that this effort will result in interdisciplinary advances that can be brought to bear on disorders of the brain itself. Correcting such disorders with treatment in their own electrical language and by individually addressing a larger set of neurons in brain circuits could be the approach that proves commensurate with the body's most complex of organs. In the long run, it could be the most revolutionary aspect of electroceuticals.

Critics will argue that we underestimate the complexity of the nervous system; the challenges in reliably, durably and non-destructively manipulating groups of individual neurons and the sheer volume of neural information flowing through these circuits. We would argue that miniaturization and big-data handling have been among the most rapidly advancing areas of scientific research in the past decade. Starting off with peripheral intervention points and simpler circuits should also help.

There are a few noteworthy unknowns,

IT'S ELECTRIC

Electroceuticals deliver electrical impulses targeting the neural circuits that regulate the body's organs and functions.



but these will be resolved only when the approach is put to test. To what extent does mapping of the neural language in animal models translate to the human setting? In which diseases will modulation of the relevant neural circuits suffice to reverse or control disease progression? Could the degree of circuit redundancy or plasticity limit the efficacy of treatment?

CATALYSING THE FIELD

At GlaxoSmithKline (GSK) and in academia, we are confident that this field will deliver real medicines, and we are mobilizing resources for this journey. This summer, the University of Pennsylvania will open its Center for Neuroengineering and Therapeutics, which will bring together researchers in medicine, engineering and business. University investigators (including B.L.) are already mapping neural circuits in humans and in cats, dogs, rodents and other models of disease. They are also building and deploying devices that modulate circuits at the neuronal level, using cloud computing to mine 'big' neural data and translating these technologies for use in tools such as antiseizure devices.

At the Feinstein Institute for Medical Research (where K.J.T. is president), scientists are trying to establish the neural codes that underlie diseases of immunity and inflammation, identify intervention points and conduct exploratory clinical work. Results so far indicate that it is feasible to identify and manipulate neural signals

specific to different inflammatory mediators in standard laboratory models.

At the Massachusetts Institute of Technology, researchers (including E.S.B.) are collaborating to map and modulate neural circuits using technologies that range from optogenetics⁶ to scalable, automated electrophysiology¹⁰ — and they are distributing the genetic codes, hardware and software necessary to put these inventions into practice.

At GSK, we (K.F. and M.S.) are committed to acting as a catalyst for this emerging field, through three immediate steps. The first, a programme that will fully fund up to 40 researchers in up to 20 external labs conducting exploratory work mapping disease-associated neural circuits, launches this week (www.gsk.com/bioelectronics). Funding for the first year will be awarded after a rapid review and approval process that should take roughly one month. Early findings will be shared among researchers in this network, and intellectual-property rights will remain with the inventors. Throughout this exploratory phase, the network will be encouraged to shape longer-term efforts in research and development.

In December, as a second step, GSK will hold a global forum for research leaders to chart an integrated path forward and to collectively identify a key hurdle in the field.

"Could the degree of circuit redundancy or plasticity limit the efficacy of treatment?"

After the forum, the company will launch the third step: a US\$1-million prize for innovation, to be awarded to the group that overcomes this hurdle.

Clearly, open innovation and flexibility in dealing with intellectual property will be important. As the poet Cesare Pavese said: "If you wish to travel far and fast, travel light. Take off all your envies, jealousies, unforgiveness, selfishness and fears." Together we can bring about the era of electroceuticals. ■

Kristoffer Famm is vice-president of bioelectronics research and development at GlaxoSmithKline, Brentford TW8 9GS, UK.

Brian Litt, Kevin J. Tracey, Edward S. Boyden, Moncef Slaoui.
e-mail: kristoffer.h.famm@gsk.com

1. Kandel, E. R., Schwartz, J. S., Jessell, T. M., Siegelbaum, S. A. & Hudspeth, A. J. *Principles of Neural Science* 5th edn (McGraw Hill Professional, 2012).
2. Koopman, F. A. et al. *Arthritis Rheum.* **64** (suppl.), abstr. 451 (2012).
3. Aponte, Y., Atasoy, D. & Sternson, S. M. *Nature Neurosci.* **14**, 351–355 (2011).
4. Huber, D. et al. *Nature* **451**, 61–64 (2008).
5. Andersson, U. & Tracey, K. J. *J. Exp. Med.* **209**, 1057–1068 (2012).
6. Chow, B. Y. & Boyden, E. S. *Sci. Transl. Med.* **5**, 177ps5 (2013).
7. Rapoport, B. I., Turicchia, L., Wattanapanitch, W., Davidson, T. J. & Sarpeshkar, R. *PLoS ONE* **7**, 34292 (2012).
8. Wang, Z. L. & Wu, W. *Angew. Chem. Int. Ed.* **51**, 11700–11721 (2012).
9. Alivisatos, A. P. et al. *ACS Nano* **7**, 1850–1866 (2013).
10. Kodandaramaiah, S. B., Franzesi, G. T., Chow, B. Y., Boyden, E. S. & Forest, C. R. *Nature Meth.* **9**, 585–587 (2012).



Evolution's red-hot radical

Sidekick status does Alfred Russel Wallace an injustice. He was a visionary scientist in his own right, a daring explorer and a passionate socialist, argues **Andrew Berry**.

Alfred Russel Wallace is too often remembered as little more than Charles Darwin's goad. Darwin's procrastination in publishing his ideas on evolution by natural selection was ended by the arrival from Indonesia of a manuscript from Wallace that outlined the very same idea. Papers by both men were read at the Linnean Society on 1 July 1858 (ref. 1). *On the Origin of Species*, published by Darwin the following year, brought the theory into the public eye.

Wallace's finest moment has condemned him to be forever Watson to Darwin's Holmes. Characteristically, Wallace actively promoted this perception. He entitled his major 1889 book on evolution *Darwinism* and, at the 1908 event to mark the 50th anniversary of the joint reading, played down his contribution: "I was then (as often since) the 'young man in a hurry': he [Darwin] the painstaking and patient student, seeking

ever the full demonstration of the truth that he had discovered, rather than to achieve immediate personal fame."

Sidekick status does Wallace an injustice. He was a visionary scientist in his own right, a daring explorer and a passionate socialist. This year's conferences and exhibitions marking a century since his death in 1913 (see go.nature.com/icpkp8) provide an excellent opportunity to reappraise his huge scientific legacy, which ranged from discovering natural selection to defining the term species, and from founding the field of evolutionary biogeography to pioneering the study of comparative natural history.

AMAZON APPRENTICESHIP

Born in 1823 into genteel poverty, Wallace left school at 13 years old to assist his brother, a land surveyor. Tramping the English countryside introduced Wallace to his first scientific interest: plants. He became a

serious student of natural history in 1844, when another young, self-educated naturalist, Henry Walter Bates (of future Batesian mimicry fame), introduced him to beetle collecting. In 1847, dissatisfied with "a mere local collection", Wallace wrote to Bates, "I should like to take some one family, to study thoroughly — principally with a view to the theory of the origin of species."

So, with extraordinary daring, the two neophytes headed to the Brazilian Amazon in 1848. Wallace stayed for 4 years, Bates 11. They funded their scientific expedition by selling specimens.

Wallace headed home in 1852. Because of a customs issue, he found many of the specimens he had been sending to London held up in Manaus in Brazil, at the confluence of the Amazon and Rio Negro. Reunited with the fruits of years of perilous labour and accompanied by a small menagerie of living animals he had ferried across the continent

to Belém at the mouth of the Amazon, Wallace must have fantasized about the impact of his arrival on London society: imagine walking into Victorian scientific salon with a toucan on your arm.

It was not to be. In the middle of the Atlantic, Wallace's ship caught fire and went up like a tinderbox. He had time to grab only a small box of drawings before being a part of what is surely one of the most poignant scenes in the history of science. Hopeful that the burning wreck would attract other shipping, Wallace and the crew stayed close. He watched as the living animals he had brought such a distance — his pets and his passport to the scientific big time in London — perished on the stricken vessel. "Many of the parrots, monkeys, and other animals we had on board, were already burnt or suffocated; but several had retreated to the bowsprit out of reach of the flames ... quite unconscious of the fate that awaited them."

Wallace spent ten days adrift in an open boat before being rescued. His Victorian stiff upper lip never quivered: "During the night I saw several meteors, and in fact could not be in a better position for observing them, than lying on my back in a small boat in the middle of the Atlantic."

Having lost almost everything but determined to make his name as a naturalist-scientist, Wallace took to sea once more in 1854, en route to Singapore, from where launched his second set of extraordinary explorations. The Amazon was his scientific apprenticeship; his eight-year journey through southeast Asia was, Wallace wrote, "the central and controlling incident of my life".

Wallace ranged from peninsula Malaysia to Western New Guinea (see page 165). This time, despite his many dangerous adventures in small boats in remote archipelagoes, Wallace's extensive collections made it safely back to England, complete with some 1,000 species new to science. By his return in 1862, he was a member of the scientific elite. He had impressed hungry collectors and institutions such as the British Museum with his constant stream of specimens. And, thanks to the Linnean Society reading and a stream of innovative papers from the field, he had made a name for himself as biological theorizer.

SCIENTIFIC DEBUT

In forcing him to undertake a second expedition, Wallace's mid-Atlantic catastrophe inadvertently completed his biogeographical education. As a collector, he was interested in the distribution of animals: he needed information on where he could find particular species, and he was sensitive to geographical transitions from one species to another. In one of his early Amazon papers, he complains about the lack of precision of previous naturalists in designating the ranges of monkey species.

Wallace had a prodigious ability to spot patterns in the apparently chaotic (and largely uncatalogued) world of tropical diversity. This is the skill of the true naturalist: to generate a mental database of observed plants and animals that can be referenced when similar forms are encountered elsewhere. It led to his first attempt at biological generalization, a paper he wrote in 1855 while in Sarawak, Borneo: "On the Law which has Regulated the Introduction of New Species" (often called the Sarawak Law)².

It was a stunning scientific debut. He italicized the paper's take-home message, "Every

"The origin of species was, Wallace recognized, a genealogical process."

species has come into existence coincident both in space and time with a pre-existing closely allied species". In other words, related species tend to be found in the same geographical area (all kangaroos are in Australasia, for instance) and, as fossils, in contiguous strata (all ceratopsid dinosaurs appear in the late Cretaceous). The origin of species was, Wallace recognized, a genealogical process.

A remarkable feature of the Sarawak law is Wallace's synthetic use of published information. In Borneo, Wallace had little or no access to fossil material: wet tropical environments are famously fossil-poor because vegetation and soil conceal underlying rock formations. In the absence of academic libraries, he relied on his phenomenal memory and whatever published accounts of the fossil record he could carry. Charles Lyell's *Principles of Geology* (1830–33) was his bible. Whereas Darwin's appreciation of the fossil record was acquired in the field in South America, Wallace's palaeontological insights were book-learned. This makes his magisterial two-volume overview of the global distribution of life — *The Geographical Distribution of Animals* (1876) — all the more extraordinary in its synthesis of living and fossil data.

Wallace's pattern-spotting led him to another concept fundamental to evolution. In his brilliant 1865 paper on the papilionid butterflies of southeast Asia³, he parses variations within and among populations, among subspecies and species, and arrives at this definition: "Species are merely those strongly marked races or local forms which, when in contact, do not intermix, and when inhabiting distinct areas are generally believed to have had a separate origin, and to be incapable of producing a fertile hybrid offspring."

It is emblematic of history's neglect of Wallace that most undergraduates today are taught that the biological species concept was introduced in 1942 by Ernst Mayr⁴.

Becoming familiar with the wet tropics of both the New and Old Worlds put Wallace

in a position to ask questions on a larger scale. Given climatic similarities, why do two regions have markedly different fauna? Wallace had laid the groundwork for this field — historical biogeography — in his 1857 paper about the Aru Islands off western New Guinea⁵, and eventually made it his own. He wrote "how totally the productions of New Guinea differ from those of the Western Islands of the Archipelago, say Borneo", despite the likeness of their "climate and physical features". He also pointed out that, despite the contrast between the physical conditions of Australia and New Guinea, "the faunas of the two, though mostly distinct in species, are strikingly similar in character." Had Borneo and New Guinea been geologically connected, Wallace hinted, their faunas would have been similar.

The Sarawak law built on this, in noting that the distributions of species are dictated partly by environmental considerations (some trees, for example, are tropical specialists) but mostly by the quirks of history. It is this line of thinking that culminated in *The Geographical Distribution of Animals*. If only Wallace had lived to see the unveiling of the theory of continental drift in the 1960s that clinched it: Australia and New Guinea are on one tectonic plate, Borneo on another.

LUCKY BREAK

It is tempting to see echoes between Wallace's serendipitous path through life and his contingent interpretation of natural systems: his most famous biogeographical discovery also had a dose of luck. In 1856, having missed a connection as he tried to make his way to Sulawesi, he spent a couple of months on the islands of Bali and Lombok, and noted drastic differences in the wildlife even though the islands are only some 35 kilometres apart. To the south and east, the Australian fauna dominated; to the north and west, the Asian one. He had identified an ancient biogeographic split across southeast Asia that biologist Thomas Henry Huxley later dubbed 'Wallace's Line'.

Wallace's 1858 discovery of natural selection pulled these strands together. While in a fever on the Maluku islands (Moluccas), he was pondering another biogeographic discontinuity: that between the Austronesian people of southeast Asia and the Melanesians of New Guinea. Drawing, like Darwin, on the work of economist Thomas Malthus, he focused on competition for limited resources. Combining this with his appreciation of variation within species that came from being a collector, natural selection was, for him, a logical step.

Wallace was disappointed that his heretical Sarawak law paper had barely made a ripple. His agent in London complained that "theorizing" was not useful and that Wallace ►



One of the few Amazon drawings rescued by Alfred Russel Wallace from his burning ship in 1852.

▶ should “collect more facts”. To save his natural-selection paper from the same fate, Wallace sent the manuscript to a senior colleague in the hope that his endorsement would give it prominence. That colleague was Darwin. How different things may have turned out had Wallace instead sent the manuscript directly to a journal.

BEYOND EVOLUTION

Standard histories of Wallace sometimes refer to the “other Wallace”, giving the impression of a flake who used his newfound scientific celebrity to plunge into dubious causes, from suffrage and socialism to spiritualism and phrenology. But Wallace’s world view was far more coherent⁶ than is often claimed. Take, for example, his perspective on human evolution.

Of the several disagreements between Wallace and Darwin, the most significant was on human evolution: Wallace came to believe that natural selection alone could not account for our species. Darwin was horrified, writing to his friend in 1869: “I hope you have not murdered too completely your own and my child.” Two factors are behind Wallace’s defection on this issue.

First, he had become a convinced spiritualist. Seances by fraudulent mediums keen to fleece fashionable Victorians were especially popular among free-thinkers such as Wallace. They had disavowed established religion but hankered for something to fill the void. Wallace even attempted to convince his scientific colleagues that spiritual forces were undetectable by scientific means because the technology had not yet been devised. In

the days before microscopes, he wrote, who would have believed that a drop of water from the Thames was swarming with tiny creatures? As a spiritualist, he had to assert the existence of some kind of non-material intervention in the genesis of humans.

His other reason for rejecting natural selection as sufficient for human evolution is more scientific. Having spent some 12 years living among and being dependent on people deemed by Victorians to be savages, Wallace, unlike even the most liberal abolitionists, was no racist. “The more I see of uncivilized people, the better I think of human nature on the whole, and the essential differences between so-called civilized and savage man seem to disappear.”

For Wallace, this enlightened social perspective posed an evolutionary problem. He appreciated that an Aru islander living in a mud hut has the same mental attributes as a member of London’s scholarly Athenaeum club. Given the necessary training, he felt, the islander would be able to play Chopin and to declaim Ovid; yet this potential would never be realized on the Aru Islands. Thus, many humans have abilities that they never have the opportunity to use.

Such a situation, Wallace reasoned, cannot evolve through natural selection alone, which promotes only those traits that are useful. Wallace concluded that human evolution required some divine intervention. This argument shows an excellent appreciation of the mechanics of natural selection, even if we are now comfortable with the idea of the brain having evolved under natural selection for specific adaptive purposes, with

many of its attributes — including Chopin-playing and Ovid-declaming — being simply by-products of the resulting organ.

Whatever one’s qualms about some of Wallace’s non-scientific causes, one cannot fail to be impressed by the passion and intensity he brought to them. He was in many ways the prototype of a socially engaged scientist. A constant theme of his 20 books and almost 800 articles is sympathy for the underdog: whether that be the poor (“To allow one child to be born a millionaire and another a pauper is a crime”); disenfranchised women (“women are human beings; therefore they should have votes as well as men.”); or the threatened redwoods of California (“Let us hope that ... care will be taken, before it is too late, to preserve ... some more extensive tracts of forest.”).

Meanwhile, Wallace remained engaged and productive as a scientist throughout his life. One of his last books, *Is Mars Habitable?* (Macmillan, 1907), arguably established the field of astrobiology (see U. Kutschera *Nature* **489**, 208; 2012). He wrote extensively on the evolution of animal colouration, especially crypsis (camouflage), aposematism (warning colouration) and mimicry. And he suggested that natural selection may actually facilitate speciation by promoting the evolution of inviability or infertility of hybrids between incipient species (sometimes called the Wallace effect⁷). The most significant of Wallace’s contributions, however, were his synthetic works on evolutionary biogeography: *The Geographical Distribution of Animals* and *Island Life* (1880), which established the field and set the bar high for future contributions.

As we remember Wallace 100 years after his death, let us celebrate his remarkable scientific achievements and his willingness to take risks and to advocate passionately for what he believed in. He was, after all, both a scientist, and, in his own assessment, a “Red-hot Radical, Land Nationaliser, Socialist, Anti-Militarist, etc., etc., etc.” In short, a whole lot more than Darwin’s good. ■

Andrew Berry is lecturer in Organismic and Evolutionary Biology at Harvard University, Cambridge, Massachusetts, USA.
e-mail: berry@oeb.harvard.edu

1. Berry, A. & Browne, J. *Nature* **453**, 1188–1190 (2008).
2. Wallace, A. R. *Annals and Magazine of Natural History* **16**, 184–196 (1855).
3. Wallace, A. R. *Transactions of the Linnean Society of London* **25**, 1–71 (1865).
4. Mayr, E. *Systematics and the Origin of Species* (Columbia Univ. Press, 1942).
5. Wallace, A. R. *Annals and Magazine of Natural History* **20**, (Suppl.) 473–485 (1857).
6. Fichman, M. *An Elusive Victorian: The Evolution of Alfred Russel Wallace* (Univ. Chicago Press, 2004).
7. Johnson, N. A. in *Natural Selection and Beyond: The Intellectual Legacy of Alfred Russel Wallace* (eds Smith, C. H. & Beccaloni, G.) 114–124 (Oxford Univ. Press, 2008).

IN RETROSPECT

The Malay Archipelago

David Quammen re-enters the ‘Milky Way of land masses’ evoked by Alfred Russel Wallace’s masterpiece of biogeography.

The Malay Archipelago: the land of the orang-utan, and the bird of paradise; a narrative of travel, with studies of man and nature

ALFRED RUSSEL WALLACE

Macmillan/Harper Brothers: first published 1869.

Alfred Russel Wallace was arguably the greatest field biologist of the nineteenth century. He played a leading part in the founding of both evolutionary theory and biogeography (see page 162). He was also, at times, a fine writer. The best of his literary side is on show in his 1869 classic, *The Malay Archipelago*, a wondrous book of travel and adventure that wears its deeper significance lightly.

The Malay Archipelago is the vast chain of islands stretching eastward from Sumatra for more than 6,000 kilometres. Most of it now falls within the sovereignties of Malaysia and Indonesia. In Wallace’s time, it was a world apart, a great Milky Way of land masses and seas and straits, little explored by Europeans, sparsely populated by peoples of diverse cultures, and harbouring countless species of unknown plant and animal in dense tropical forests. Some parts, such as the Aru group of islands, just off the coast of New Guinea, were almost legendary for their remoteness and biological riches. Wallace’s journeys throughout this region, sometimes by mail packet ship, sometimes in a trading vessel or a small outrigger canoe, were driven by a purpose: to collect animal specimens that might help to answer a scientific question. That question was: if species evolve, what is the mechanism?

His Malay expedition began in 1854, five years before the publication of Charles Darwin’s *On the Origin of Species*. As he pursued his goal, oblivious to Darwin’s slow, secret theorizing, Wallace suffered one disadvantage that, paradoxically, proved advantageous: a need to earn money. He had no family funds such as those that had eased Darwin’s way. He paid his expenses by selling select specimens — pretty beetles and butterflies, stuffed birds and occasionally mammal skins — to museums and amateur collectors through an agent in London. So he had reason to take multiple individuals of the more striking species. He collected series, not just samples. From a riverside in southern Celebes he brought away “six good specimens” of *Papilio androcles*, a rare and beautiful swallowtail butterfly. And in Waigiou he harvested “twenty-four fine specimens” of *Paradisaea rubra*, the red bird of paradise. One effect of this redundant, commercial collecting was that he saw intraspecific variation laid out before him. ▶

“Wallace paid his expenses by selling specimens. So he collected series, not just samples.”



Birds of paradise, as illustrated in *The Malay Archipelago* in 1869.

ERNST MAYR LIB., MUS. COMPARATIVE ZOOLOGY, HARVARD UNIV.

► At a time when essentialist thinking (every species fixed, an ideal type) and the idea of special creations (each species shaped in a particular way by God) prevailed, even among most scientific naturalists, to see intraspecific variation was to take the first step towards an evolutionary theory.

Wallace travelled continuously among the islands for eight years (until 1862), and his mishaps, his hardships, his long weeks of illness and loneliness, his near-death scrapes, as well as his collecting and observing, are recounted in this vivid, encompassing book. From his jumping-off point in Singapore he proceeded to Borneo, then to Bali, then across the narrow, deep strait separating Bali from Lombok. From there he looped up to the Dutch entrepôt of Macassar, on the southwest tip of Celebes, which became his hub for further voyages around the archipelago.

He zigzagged from island to island — out east to the Aru cluster, back west to Sumatra and Java, up into the northern Moluccas and down south to Timor — depending

on his own whims and interests, but also on which way the monsoon winds were blowing or the next mail ship was heading. He took note of people and cultures, and recorded what intrigued him in his affable, bemused voice. Those parts of his book are as chatty and engaging as Samuel Pepys's diaries or James Boswell's *Life of Johnson*. And he continued his biological collections: an exquisite birdwing butterfly here, a giant beetle there, and, whenever possible, multiple specimens of each.

When the time came to combine his collecting notes with his anthropological observations and the diary of his many adventures and misadventures, Wallace recast the zig-zaggy muddle into a linear geographical structure. *The Malay Archipelago* proceeds from west to east — Singapore to Aru.

This structure serves well the implicit subjects of the book: biogeography and

"This book is a joyride through one of the most exotic, most remote regions of Earth."

evolution. Comparing the fauna of Bali with that of Lombok, Wallace laid the groundwork for one of his major contributions: the concept of Wallace's Line, the deep-water division between two major faunal realms, the Indo-Malay and the Australasian. Moving from Sumatra to Java to Borneo, he suggested (correctly) that those islands must once have been connected to the Asian mainland, resting high and dry on what we now call the Sunda Shelf. And he supported Darwinian theory, subtly rather than explicitly, by showing, in case after case, that only dispersal of ancestral forms, followed by isolation and evolutionary change, could explain the patterns of faunal distribution that his long labours had revealed. Special creations and essentialism just couldn't make sense of the Malay Archipelago.

But this book is not just a subtle compilation of data and argument. It is also a joyride through one of the wildest, most exotic, most remote regions of Earth that any Victorian explorer ever visited. There are deft descriptions of sensory experience, such as this on the taste of durian, the East's most notorious fruit: "A rich butter-like custard highly flavoured with almonds ... but intermingled with it come wafts of flavour that call to mind cream-cheese, onion-sauce, brown sherry, and other incongruities." There are moments of breathless drama, mostly involving not physical danger (about which Wallace tends to be matter-of-fact) but small, important triumphs, such as when he caught a new species of birdwing butterfly: "On taking it out of my net and opening the glorious wings, my heart began to beat violently, the blood rushed to my head, and I felt much more like fainting than I have done when in apprehension of immediate death." From his high points of discovery to his low points of misery, and through all the miles and days in between, Wallace is a companionable narrator with a dry wit, a keen eye, an inexhaustible curiosity and not a trace of self-pity. What more can readers ask of a literary and scientific traveller?

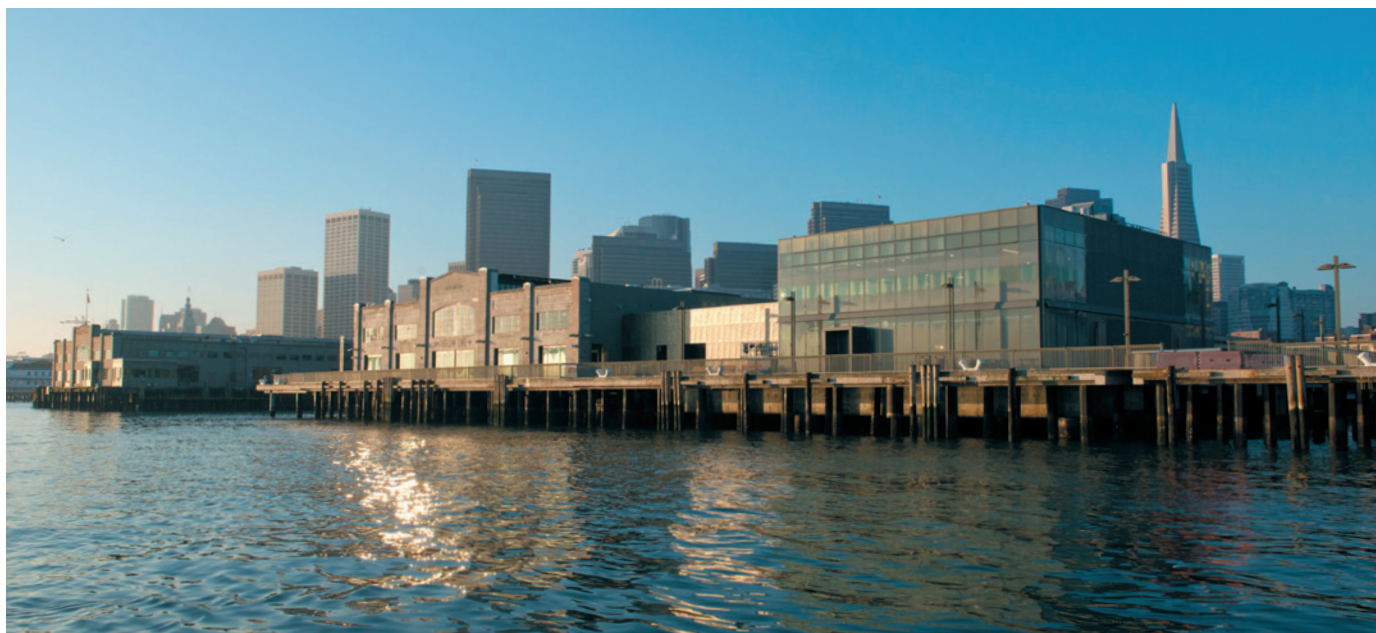
One thing is curiously absent from this long, brimming book. In recounting his stay on Ternate, in 1858, Wallace declined even to mention the world-altering idea — his theory of evolution — that he put to paper there. He could have crowed: in this place, at this time, I co-discovered natural selection. Evidently he didn't feel the need. By 1869, when *The Malay Archipelago* appeared, Wallace himself was a Darwinist. ■

David Quammen is a writer. His most recent book is *Spillover: Animal Infections and the Next Human Pandemic*. His 1996 book, *The Song of the Dodo: Island Biogeography in an Age of Extinctions*, features Wallace.

e-mail: quammen@imt.net



Some of the tropical butterflies that Wallace collected during his travels.



SCIENCE EDUCATION

A science giant moves house

Lucy Odling-Smee scopes out the new waterfront abode of Frank Oppenheimer's San Francisco museum.

As you approach Pier 15 on San Francisco's waterfront, it's easy to imagine that the giant stone facade conceals another of the area's many car parks. Yet inside this former bus-repair station, hundreds of contractors in yellow hats and scurrying curators are shaping a scientific wonderland of gadgetry and natural phenomena.

San Francisco's Exploratorium museum is moving home. Next week, following a US\$220-million construction and relocation project, it will re-open, nearly half a century after physicist Frank Oppenheimer — brother of J. Robert Oppenheimer — founded it.

Now hovering above the Bay, the museum has three times the space it had previously, at the cavernous Palace of Fine Arts four kilometres down the road. And whereas the Palace could be gloomy, here the Californian sun streams in through 5-metre-high windows.

After being invited to tinker with the museum's surprisingly addictive classic exhibits — about motion, magnetism and light — visitors will enter six new life- and environmental-sciences galleries. Here they can encounter anything from developing

zebrafish eggs to dancing plankton.

The aim of many of the new exhibits is to reveal the extraordinary in the ordinary. "When it's out of context, you see it," says

Kristina Yu, director of the Living Systems Department and co-curator of the East Gallery. Take trees, for example. It's easy to pass them by. Yet the enormity of a nine-tonne section of a Douglas fir, lying prostrate in one corner of the East Gallery, demands your gaze. Close up, visitors can learn about the history of the tree's life, recorded in its rings.

The tidal drama of the Bay — little noticed outside — is revealed in what looks like a giant Peruvian flute: an arc of 24 clear plastic columns, more than 3 metres high and part-filled with water. The water level in the tubes, each of which represents one hour on the clock, is controlled by air pressure to move in synchrony with the seawater below the pier. When a tube's given hour ends, the level is frozen, producing a record of daily ebbs and flows.

Exploratorium

PIER 15, SAN FRANCISCO, CALIFORNIA
Grand Opening of the new site, 9 am to 10 pm on 17 April 2013

That "the space itself has led the ideas", as Yu explains, is also obvious in the Fisher Bay Observatory: a two-storey glass building with spectacular views, one side facing the water, the other the glitter of downtown San Francisco. Outside, sailing boats dart between bulkier vessels. On a nearby screen, meaning emerges from six months of data on boat routes. Slow-moving blue, green or pink arrows indicating the passage of an individual ferry, tanker or tugboat become rivers of colour as you turn a dial to track all the journeys that took place during the past hours, weeks or months.

Other data visualizations are projected onto a topographic three-dimensional map of the Bay Area. More button pressing, and impressively realistic fog gathers and crumbles. A rash of circles rippling outwards shows where earthquakes have struck. Scientists from the US National Oceanic and Atmospheric Administration are helping the curators to gather real-time data on air quality, the weather, tides and water pollution. The hope is that the data will spawn other displays — and, as part of an observational network, the Exploratorium will aid researchers in return.

I was sceptical of what I might find at the relocated Exploratorium. The wealth of moving parts, buttons and levers in many science museums often makes me feel like I am in a playground after the school day finishes, with no sky above my head and no way out. But, so far, everything at Oppenheimer's new-old museum indicates that the curators have mastered not just the art of engagement, but how to show people the most interesting stuff of all — what is right under their noses. ■

Lucy Odling-Smee is a Nature editor based in San Francisco, California.



Q&A Genevieve Dion

The digital knitter

Genevieve Dion works at textile engineering's cutting edge at Drexel University in Philadelphia, Pennsylvania. Ahead of the Smart Fabrics conference in San Francisco, California, she talks about knitting robots, permanently pleating silk and charging mobile phones from shirts.



How did you start out in fashion?

Studying industrial design in San Francisco, I took a class in *shibori*, a Japanese technique where you bunch up a cloth and soak it with colour, like tie-dye. I had to develop my own equipment to fold the fabric so it resisted and absorbed dye in particular geometric configurations. This led me to explore fabric sculpting and seamless garments using woven silk from Japan. Seeking new modes of manufacturing became almost more interesting than the clothing itself.

How did you begin to pleat silk?

There are many industrial machines that produce perfectly pleated textiles such as the ones used for Issey Miyake's Pleats Please range of clothes. Inspired by the *shibori* process, I developed my own method of chemically scoring silk so that it retains texture without stitches. It is a trade secret, and labour-intensive, like giving silk a perm. Unlike the silk-pleating technique used by early

twentieth-century fashion designer Mariano Fortuny, our method is permanent. If the pattern relaxes, you throw the dress or scarf in the dryer and it comes out pleated again.

What inspired you to make 'smart garments'?

Haute-couture designers such as Hussein Chalayan along with some crafty do-it-yourself designers have incorporated electronic components into their clothes, turning them into expressive wearables that transmit, emit and communicate. But mass production hasn't been fully worked out. Issues such as the integration of flexible circuitry and adequate power sources have to be resolved to make wearable technology reliable and marketable. I wanted to rethink smart garments from scratch.

Tell me about smart garments for health care.

I came to Drexel for its fashion programme, but its medical and engineering programmes also wanted to design smart garments. I now work with a team of engineers and scientists. We have learned to speak a common language, although sometimes I

just have to make a prototype and give it to the engineers to analyse. The garments we are working towards will include wearable sensors to measure breathing, heart rate and other vital signs for easy monitoring of patients, or sensors that can detect potential hazards such as obstacles or harmful chemicals. The sensors will have to be almost invisible to allow people wearing them to blend in. This technology has the potential to convey critical information, including written and spoken reminders to a patient. We are also working on carbon-coated, conductive textiles for flexible energy storage, which might allow wearers to charge their cellphone from their shirt.

And you've taken up industrial knitting?

Knitting is versatile, and machines are already knitting clothes on an industrial scale. With a single piece of string you can make a garment in any shape. You can do rapid prototyping. There are carbon-fibre yarns, stainless-steel yarns, coated yarns, yarns incorporating LEDs: if we can knit the high-tech yarns into garments, we can mass produce these. And because knitted garments are made one at a time, there is the possibility of mass customization. A mechanical understanding is necessary, as with any craft, but the software gives you the power to visualize the resulting garment before it is made.

What is next for you?

I'm working with engineers to replace the metal and plastic shells of robots with protective, abrasion-resistant knitted fabrics. We call it 'soft robotics'. Knitting could also be a new form of industrial fabrication for internal components for machines. If my machine can make a glove with five fingers, or a dress with pleats and sleeves, then I can make many topological shapes with that same equipment. I did think at some point, wouldn't it be nice to knit heart valves? But when I saw that people were experimenting with 3D bioprinting them, I gave up the idea.

Do you miss making clothes by hand?

In the past I enjoyed making unique, intricate garments and accessories by hand, including commissions for the performers Tina Turner and Elvis Costello. Working with a knitting machine is a similar process, just using different tools. Despite automation it is still not possible to produce good work without a thorough understanding of the equipment and a feel for the materials. If we want, we can make limited runs of a given textile or colour, or engineer 'one of a kindness' into each garment's manufacture. So we can find new ways of using machines to offer the exclusiveness of handmade work. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

Antibiotics: avert an impending crisis

Bacterial pathogens are rapidly becoming resistant to almost all antibiotics (*Nature* **495**, 141; 2013). To make matters worse, the drug-discovery pipeline of new antimicrobials is almost empty (M. S. Butler and M. A. Cooper *Curr. Drug Targets* **13**, 373–387; 2012). The major funding bodies therefore need to give top priority to research into antibiotic discovery and resistance.

In a snapshot survey last month of the websites of 11 of the world's largest medical and scientific research-funding agencies (in the United States, Europe, Canada and Australia), I found none that allocated substantial grant money to antibiotic discovery or development, or to research on antibiotic resistance.

The funding agencies I have surveyed might argue that properly constructed applications all receive a fair hearing, but that is not the same thing. Researchers are generally adept at tailoring their grant proposals to fit funding priorities, particularly when prompted by a large financial carrot.

Peter Speck *Flinders University, Adelaide, Australia.*
peter.speck@flinders.edu.au

Antibiotics: relax UK import rule on fungi

With the UK Chief Medical Officer last month highlighting the “catastrophic threat” of antibiotic-resistant bacteria (see *Nature* **495**, 141; 2013), we suggest that the United Kingdom re-evaluates its misplaced over-regulation of importing fungi for research. This is obstructing a potentially effective route towards alleviating the antibiotics crisis.

Some of the most successful antimicrobials, among other pharmaceuticals, are derived from fungi — including penicillins, cephalosporins, cyclosporins, griseofulvins and pneumocandin. Given that only

2–12% of all estimated fungal species have been described and less than half of those in the GenBank database have been named, unknown fungi may have the greatest potential for yielding new pharmaceuticals.

As implemented, The Plant Health (England) Order of 2005 requires an import licence for all live fungi. Licensing requires the species to be named and risk-assessed, preventing researchers from using any unknown fungi.

The biggest risk to plant and human health is from fungal spores entering the country accidentally: for example, in dust and soil on vehicles, shoes, packaging, migrating birds or insects, and horticultural products. Considering the scale of this unregulated import of spores, we suggest that the constraints imposed by current policies are unrealistic as well as counterproductive.

David L. Hawksworth *Universidad Complutense de Madrid, Spain; and The Natural History Museum, London, UK.*
d.hawksworth@nhm.ac.uk
Bryn T. M. Dentinger *Aberystwyth University, Aberystwyth, UK.*

Naples fire inflames rise in creationism

On the night of 4 March, unknown arsonists destroyed the famous City of Science centre in Naples. This shocking event has been publicly welcomed by some hard-line creationists in Italy, who disapprove of the teaching of evolution at the centre.

However, the minister for education and research, along with the mayor and the regional governor, have agreed to reconstruct the buildings within two years. The founder of the City of Science, physicist Vittorio Silvestrini, announced that the centre will open new exhibition areas in the surviving buildings and gardens this week — backed by donors and

other science museums. In the week after the fire, about 10,000 people from Naples marched in protest against the arson attack.

The conservative commentators who instead seemed to rejoice at the centre's destruction include the national newspaper *Il Foglio*, which applauded the flames as purification against the scourge of evolutionism — “a nineteenth-century superstition” destined to bring young people to “despair and simian behaviours”. This inflammatory anti-Darwinism has been rapidly amplified on the Internet.

Italy's current climate of political uncertainty seems to have fostered an unexpected resurgence in creationism — with science education a prime target.

Telmo Pievani *University of Padua, Italy.*
dietelmo.pievani@unipd.it

Guaranteeing high standards takes time

My impressions from a 17-year tenure at the US National Research Council (NRC), part of the National Academy of Sciences, do not tally with Marjory Blumenthal's views of the academy's performance (*Nature* **494**, 423–424; 2013).

The NRC's investment in procedures aimed at assuring objectivity, independence and scientific rigour is impeccable. Ensuring the credibility of scientific advice takes time. The process cannot be influenced by the impatience of advocates under pressure from their clients, despite today's expectations for rapid information. It would indeed be desirable for the NRC to be more “nimble”, but not at the expense of its scientific standards.

I find it commendable that the NRC holds steady in its allegiance to fulfil its advisory role. In brushing over

the complexities of process and product, I believe that Blumenthal misjudges the pursuit of consensus as a worthy end in itself.

Michael Feuer *The George Washington University, Washington DC, USA.*
mjfeuer@email.gwu.edu

Anonymity of sperm donors under threat

Personal genomics services are becoming popular for genealogical or ancestry-tracing purposes (see, for example, go.nature.com/scbwci). But this direct-to-consumer practice undermines promises of sperm-donor anonymity, which is still common in clinics in many countries (see also M. Angrist *Nature* **494**, 7; 2013).

Adults conceived from donor sperm can now identify their biological fathers using DNA testing (see R. Lehmann-Haupt *Slate*, 1 March 2010). In doing so, they could be unwittingly infringing the gamete donor's constitutional right to privacy.

Companies selling personal genetic testing therefore have a responsibility to inform customers about the potential risk of privacy breaches, as do clinics accepting gamete donors.

Pascal Borry, Olivia Rusu *KU Leuven, Belgium.*
pascal.borry@med.kuleuven.be
Heidi C. Howard *INSERM and University of Toulouse III – Paul Sabatier, France.*

Two sentences to impress

Roberta Kwok might have sold her message “Two minutes to impress” more effectively if she had followed her own example and presented her points as a quick-fire “elevator speech” (*Nature* **494**, 137–138; 2013).

Josef Settele *Helmholtz-Centre for Environmental Research — UFZ, Halle, Germany.*
josef.settele@ufz.de

Sensory noise drives bad decisions

The philosopher Albert Camus once said, “Life is the sum of all your choices”. Work using an innovative experimental design in humans and rats shows that many of the errors in those choices come from the senses, not from cognition.

MATTHEW T. KAUFMAN
& ANNE K. CHURCHLAND

The decisions that humans and animals make define their existence. Some decisions are carefully considered, such as choosing a mate; other decisions must be quick, such as deciding whether a rustle in the bushes is a companion or a predator. In many situations like the latter one, decisions are based on imperfect sensory evidence accumulated over a second or two, so that the brain can average out background ‘noise’. Although we do impressively well in making these quick decisions, we make many mistakes. Writing in *Science*, Brunton *et al.*¹ combined computational modelling with behavioural studies of rat and human decision-making to uncover where the errors in simple decision-making creep in.

Previous work^{2,3} in humans and primates indicates that there is noise in decision-making, although the source of the noise has been unclear. This ‘bottleneck’ might be anywhere: in the information from the eyes and ears, in the ability to accumulate the incoming information over time, in arbitrary biases present even before the evidence starts to appear, or in an imperfect strategy that might incorrectly estimate the importance of strong bursts of sensory evidence.

Brunton and colleagues find that errors stem from imperfect sensory evidence, that is, from the initial transfer of these signals to the brain (Fig. 1). So it seems that sensory limitations hold us back from being able to perfectly evaluate incoming evidence.

The authors’ experiment was deceptively simple. They presented rats and humans with a precisely controlled but highly variable stimulus — two randomly generated streams of clicks, one originating from the left and the other from the right (Fig. 1). Subjects had to judge whether more of the clicks came from the left or the right. Several aspects of the experimental design allowed the researchers to tease apart where errors came from: the exact pattern and number of clicks varied in each trial; these trial parameters were known by the experimenters and could be exploited in the analysis; and an enormous database of decisions was amassed to increase the statistical power of the results.

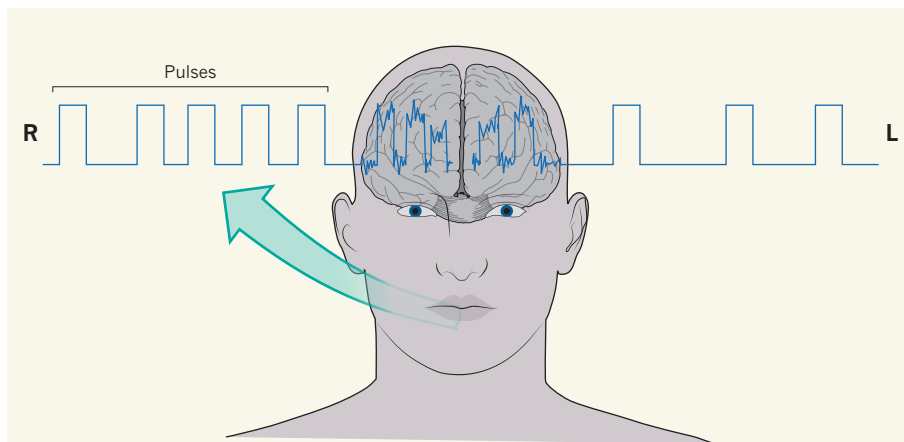


Figure 1 | Errors in decision-making. In Brunton and colleagues’ decision-making tasks¹, auditory clicks (depicted as pulses) randomly come from the left and right. As they enter the brain, the pulses must be transformed into an internal representation, a process full of opportunities for loss of fidelity. This may lead a decision-maker to incorrectly estimate the number of pulses coming from each direction. In the example shown, the subject gets it right, deciding that more clicks have come from his right.

The authors ruled out several candidate explanations for why subjects make mistakes. First, they observed that subjects performed equally well on trials that had many different patterns of clicks. This allowed the authors to exclude the possibility that errors were caused by subjects being swayed too strongly by bursts of clicks. Second, the authors evaluated decisions in response to stimuli of differing durations. Brunton *et al.* could thereby rule out two more sources for subjects’ errors: ‘muddling’ of the active memory owing to the passage of time and ‘forgetting’ clicks that came early in the trial.

The lack of muddling and forgetting is surprising; although the fundamental timescale of neuronal activity is fast, on the order of 10 milliseconds, it seems that the brain as a whole can avoid losing information over the several-second stimulus duration that defined the longest trials in the study. The long timescale of evidence accumulation points to a network of neurons that collectively keep the memory of early sensory evidence alive⁴. Noise must therefore enter earlier in the process.

The authors went one step further: they formulated a flexible mathematical model to determine each source of noise quantitatively. Although previous studies^{5,6} had concluded that the brain can reliably accumulate evidence

over time, this study is the first to simultaneously provide a quantitative estimate of the timescale involved in decision-making (at least a few seconds) and of the other sources of noise.

Brunton and co-authors mainly studied rats in this work. The rat is a relative newcomer to the field of perceptual decision-making. But although primates remain indispensable for neuroscience research because of their greater ability to be trained and their evolutionary closeness to humans, rats offer some key advantages. First is higher throughput: rat experiments can be conducted faster than primate ones. Consequently, rodent studies can include a large cohort of subjects, making it possible for experimenters to distinguish average behaviour from idiosyncratic behaviour. Second, a growing battery of molecular tools is becoming available for use in rodents, allowing easier tracing and manipulation of specific brain circuits⁷.

The present paper opens the door to a number of new directions in the study of decision-making. The ability to pin down the source of noise could reach well beyond the task for which Brunton *et al.* used it. Whereas their results suggest that most of the wrong choices the subjects made were due to errors early in the sensory pipeline, with different tasks

the errors might be driven by other sources of noise. The authors' modelling techniques could be adapted to explore the source of errors in other tasks, such as value-based decision-making^{8,9}.

An appealing possibility is that the sources of noise will be highly task-dependent. For example, memory might form the bottleneck when comparing stimuli over long periods, as when searching for a new apartment. Or perhaps pre-existing biases dominate when the cost of switching choices from one trial to another is high, as in foraging tasks in which subjects must choose whether to keep exploiting their current, but diminishing, resource or to pay a cost to explore another¹⁰. It would also be intriguing to compare the magnitude of the sensory noise in multisensory decisions¹¹, in which the presence of two sensory inputs

might increase the sensory noise.

In all types of decisions, estimation of noise parameters could be a starting point for investigating neural circuits. Many open questions remain in the field of decision-making, both in terms of the network of neural structures that are required for different kinds of decisions and the microcircuits that support computation within those structures. With the host of tools now available for dissecting neural circuits during decision-making, the coming years could bring major insights into the mechanisms underlying this crucial ability. ■

Matthew T. Kaufman and Anne K. Churchland are at Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. e-mail: churchland@cshl.edu

1. Brunton, B. W., Botvinick, M. M. & Brody, C. D. *Science* **340**, 95–98 (2013).
2. Shadlen, M. N. & Newsome, W. T. *J. Neurosci.* **18**, 3870–3896 (1998).
3. Churchland, A. K. *et al. Neuron* **69**, 818–831 (2011).
4. Wong, K.-F., Huk, A. C., Shadlen, M. N. & Wang, X.-J. *Front. Comput. Neurosci.* <http://dx.doi.org/10.3389/fncom.2007.10.006> (2007).
5. Huk, A. C. & Shadlen, M. N. *J. Neurosci.* **25**, 10420–10436 (2005).
6. Kiani, R., Hanks, T. D. & Shadlen, M. N. *J. Neurosci.* **28**, 3017–3029 (2008).
7. Callaway, E. M. *Trends Neurosci.* **28**, 196–201 (2005).
8. Kimchi, E. Y. & Laubach, M. *J. Neurosci.* **29**, 3148–3159 (2009).
9. Krajbich, I., Lu, D., Camerer, C. & Rangel, A. *Front. Psychol.* **3**, 193 (2012).
10. Adams, G. K., Watson, K. K., Pearson, J. & Platt, M. L. *Curr. Opin. Neurobiol.* **22**, 982–989 (2012).
11. Raposo, D., Sheppard, J. P., Schrater, P. R. & Churchland, A. K. *J. Neurosci.* **32**, 3726–3735 (2012).

OPTICAL DEVICES

Photonic insulators with a twist

Using a hexagonal array of helical waveguides, physicists have observed robust optical waves that move in one direction, bypassing obstacles and imperfections exactly as predicted by the theory of topological insulators. **SEE LETTER P.196**

YIDONG CHONG

Photonic researchers have long been inspired by a conceptual analogy between electromagnetic waves in patterned media and quantum-mechanical electron waves in solids. In the late 1980s, Yablonovitch¹ and John² conceived of the photonic crystal — a spatially periodic nanostructure — as an analogue of an electronic insulator. In a photonic crystal, the frequencies of propagating light can be restricted to distinct bands, which are separated by gaps of forbidden frequency, akin to the electronic bandgaps in insulators within which no electron states exist. Photonic bandgaps are now a staple of optical nanostructure design, with applications ranging from low-loss optical resonators to slow-light waveguides³. On page 196 of this issue, Rechtsman *et al.*⁴ report the design and realization of a photonic device that mimics a state of insulating matter known as a topological insulator, which has properties strikingly distinct from those of conventional insulators. This points the way towards the development of fault-tolerant photonic devices that exploit 'topologically protected' light waves.

Like a conventional insulator, a topological insulator has a bandgap. However, its bands are topologically non-trivial: they cannot be

continuously deformed into the bands of a conventional insulator as long as they remain separated by a bandgap, just as a torus cannot be transformed into a sphere without tearing its surface⁵. Consequently, when a topological insulator is connected to a conventional insulator (or a vacuum), a low-dimensional metal emerges at an edge between the two materials. Unusually, this metal is immune to conversion into an insulator by the presence of impurities or distortions on the edge, because its existence is enforced by topological differences between the materials on each side of the edge. Since the theoretical prediction⁶ and experimental discovery⁷ of topological-insulator materials in 2006 and 2007, respectively, condensed-matter physicists have been scrambling to exploit the unique properties of the materials in applications such as spintronics and fault-tolerant quantum computing.

In 2009, researchers realized⁸ a photonic system with topologically non-trivial bands using a magnetic photonic-crystal slab. (Strictly speaking, this system was analogous not to a topological insulator, but to the closely related 'quantum Hall gas'.) They showed⁸ that this photonic crystal's edge acted as a one-way waveguide, supporting scatter-free, disorder-immune wave propagation — the photonic analogue of the exotic edge metal. This system seemed to have a potential application

as an isolator, a device that permits light to flow in only one direction. But it came with two important technical shortcomings. First, the properties of the magnetic material from which the photonic crystal was made limited the operating frequency of the device to microwave frequencies rather than optical frequencies, which are of much greater technological interest. Second, the system required a strong external magnet to function, restricting its application as a stand-alone device.

Rechtsman *et al.* demonstrate a topologically non-trivial photonic system that overcomes both limitations: it operates at optical frequencies and does not require an external magnetic field. The authors' system is based on a type of topological insulator known as a Floquet topological insulator, which was proposed⁹ in 2011 and independently, in a photonic context¹⁰, in 2012. The authors of these studies noted that when a system is driven by an oscillating potential (such as an oscillating voltage), its state at each discrete period of the oscillation is equivalent to that of an undriven transformed system, which can be described using a nineteenth-century mathematical result known as Floquet's theorem. Crucially, even if the original system has topologically trivial bands, the transformed system can have topologically non-trivial bands if an appropriate driving potential is chosen.

Rechtsman *et al.* add another twist to this idea. An array of coupled parallel optical waveguides can act as a quantum-wave simulator: the flow of light down the array is formally equivalent to the evolution in time of a quantum-mechanical matter wave with one fewer spatial dimension than the array, with the distance along the waveguide axis playing the part of the time dimension. Such waveguide arrays have been used in experimental demonstrations of 'quantum-wave' phenomena, such as Anderson localization, in an entirely classical electromagnetic context¹¹. Rechtsman and colleagues go on to show that a helical

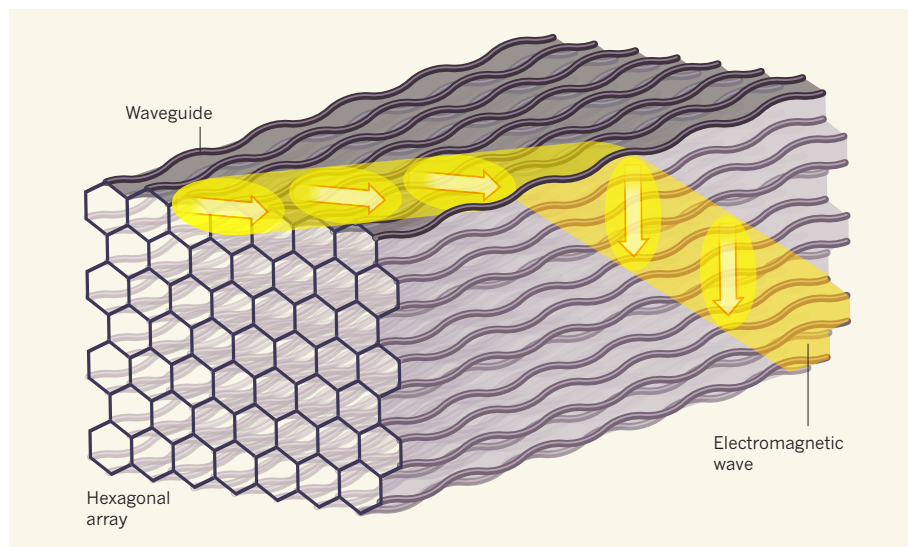


Figure 1 | A photonic analogue of a Floquet topological insulator. Rechtsman *et al.*⁴ have designed a hexagonal array of helical optical waveguides that functions as a photonic topological insulator. As an electromagnetic wave travels down the axes of the waveguides, it also moves sideways along the boundary of the array. This sideways motion is that of a 'topological edge state', and so is immune to back-scattering: on reaching the end of one edge of the array, the wave turns 90° and continues moving along the next edge. For simplicity, in this illustration the waveguides are wave-like rather than helical.

twist in the waveguides is formally equivalent to an oscillating potential, which, through Floquet's theorem, yields topologically non-trivial bands. Their experiments directly demonstrate the existence of an electromagnetic wave that moves in a single direction around the edge of the waveguide array (Fig. 1), bypassing obstacles and imperfections exactly as predicted by the theory of topological insulators.

It is worth noting that this photonic topological insulator cannot be used as an optical isolator. The waveguide array is a

three-dimensional photonic crystal composed of non-magnetic material, and a well-known principle based on the time-reversal symmetry of such systems shows that they cannot act as isolators. Thus, for each edge wave travelling in one direction along the edge of the waveguide array, it is possible to excite another wave that moves backwards along the edge and backwards along the waveguide axis. (For those familiar with models of two-dimensional electronic topological insulators, the direction of propagation along the waveguide axis in this

system can be regarded as playing the part of the electron's spin orientation.)

Although several other research groups have proposed different schemes for photonic topological insulators (see ref. 4 for references and ref. 12 for a recent experimental demonstration in an optical-chip platform), Rechtsman and colleagues' method is notable for its simplicity and practicality. The robust properties of the topological edge waves indicate several possible device applications, such as carrying signals robustly through optical fibres. Future variants of this photonic topological insulator could also be used to explore many issues of fundamental scientific interest, including how the edge waves behave under conditions of nonlinearity, amplification and damping, all of which are easily achievable and tunable in photonic media. ■

Yidong Chong is in the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore.
e-mail: yidong@ntu.edu.sg

1. Yablonovitch, E. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
2. John, S. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
3. Joannopoulos, J. D., Johnson, S. G., Winn, J. N. & Meade, R. D. *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, 2008).
4. Rechtsman, M. C. *et al. Nature* **496**, 196–200 (2013).
5. Moore, J. E. *Nature* **464**, 194–198 (2010).
6. Bernevig, B. A. *et al. Science* **314**, 1757–1761 (2006).
7. König, M. *et al. Science* **318**, 766–770 (2007).
8. Wang, Z. *et al. Nature* **461**, 772–775 (2009).
9. Lindner, N. H. *et al. Nature Phys.* **7**, 490–495 (2011).
10. Fang, K. *et al. Nature Photon.* **6**, 782–787 (2012).
11. Schwartz, T. *et al. Nature* **446**, 52–55 (2007).
12. Hafezi, M., Fan, J., Migdall, A. & Taylor, J. Preprint at <http://arxiv.org/abs/1302.2153> (2013).

and Kim *et al.*³ (page 219) address these questions using optogenetics to manipulate distinct neuronal subpopulations in mice and so dissect out the contribution of intermixed but functionally distinct cell groups*.

Both teams analysed a large, diffuse brain region called the bed nucleus of the stria terminalis (BNST). Previous studies^{4–7} have found that lesions of the BNST reduce anxiety and fear of specific environments. Other work has discovered^{8,9} distinct subregions and subpopulations of BNST neurons, and has found that the region has connections with several other brain areas that are involved in motivated behaviour and stress responses. However, the functions of the various BNST subpopulations and subregions, as well as the significance of these connections, have remained unclear.

Jennings and colleagues focused on the role of the ventral BNST (vBNST) in mediating anxiety and regulating motivated behaviour, which, along with several other behaviours, may be modulated by anxiety. Consistent

*This article and the papers under discussion^{2,3} were published online on 20 March 2013.

NEUROSCIENCE

Anxiety is the sum of its parts

Anxiety does not arise from a single neural circuit. An interplay between neighbouring, yet opposing, circuits produces anxiety, and outputs from these circuits regulate specific anxiety responses. SEE LETTERS P.219 & P.224

JOSHUA P. JOHANSEN

We all know anxiety. We might have experienced it while waiting to hear about a promotion at work, or on our way to see the doctor because she wants to talk about test results in person. A diffuse uneasiness, sometimes accompanied by perspiration and subtle changes in breathing, anxiety ebbs and flows depending on life's circumstances, and can even occur for no

apparent reason. The condition can be healthy and adaptive, but research in the United States¹ shows that, for roughly one-third of people, anxiety is a debilitating disorder at some point in their lives. Nevertheless, answers to important questions — such as how different neuronal populations represent anxiety, and how the various components of the anxious state are constructed and represented in neural circuits — remain elusive. In two studies in this issue, Jennings *et al.*² (page 224)

with the idea that the vBNST contains functionally distinct cell populations, the authors found that learned anxiety that is associated with specific environments leads to increased activity of some vBNST neurons and decreased activity of others.

Both of these cell populations made specific synaptic connections with neurons of another brain region called the ventral tegmental area (VTA), which is known to guide motivated behaviour. Specifically, cells that were excited by anxiety-inducing environments in turn excited their VTA partner, and stimulating these excitatory vBNST–VTA connections increased anxiety and decreased reward-seeking behaviour. By contrast, vBNST neurons that were inhibited by anxiety-inducing environments also inhibited their downstream VTA neurons, and stimulating these inhibitory connections promoted reward-seeking behaviour and reduced anxiety.

A caveat of this work is that the authors did not inhibit vBNST–VTA connections during natural anxiety states, but rather stimulated the neurons to regulate anxiety and motivated behaviours. Thus, it is possible that engagement of these circuits by anxiety does not produce the same behavioural effects naturally as those seen with artificial stimulation. However, the fact that during anxious states the vBNST–VTA neurons, which are known to promote anxiety, were activated and those that reduce anxiety were inhibited provides strong correlative evidence that learned anxiety naturally engages these neuronal subpopulations. The interplay between these two opposing ‘push–pull’ circuits may set an adaptive, or even a maladaptive, level of anxiety, and allow for bidirectional regulation of reward-motivated behaviour during anxiety.

Kim and co-workers asked whether, and if so how, cells in the two subregions of the dorsal BNST, the oval nucleus (ovBNST) and the anterodorsal BNST (adBNST), differentially regulate anxiety. They found that the activity of ovBNST neurons promoted anxiety. Moreover, inputs from the amygdala, a brain region that has been implicated in fear, reward and anxiety, activated adBNST neurons and reduced anxiety, and inhibition of these inputs increased anxiety. Consistent with a role in reducing anxiety, adBNST neurons fired more when the animals were in a safe environment than when they were in an anxiety-producing one, thus distinguishing between the two places (Fig. 1).

Intriguingly, inhibiting amygdala inputs to the adBNST reduced the ability of this subregion’s neurons to distinguish between safe and anxiety-producing places, which suggests that adBNST cells reduce anxiety in response to a ‘safety’ signal from the amygdala. Future work should determine how amygdala neurons connecting to the adBNST encode anxiety-related information and what types of experience recruit this anxiety-reducing circuit.

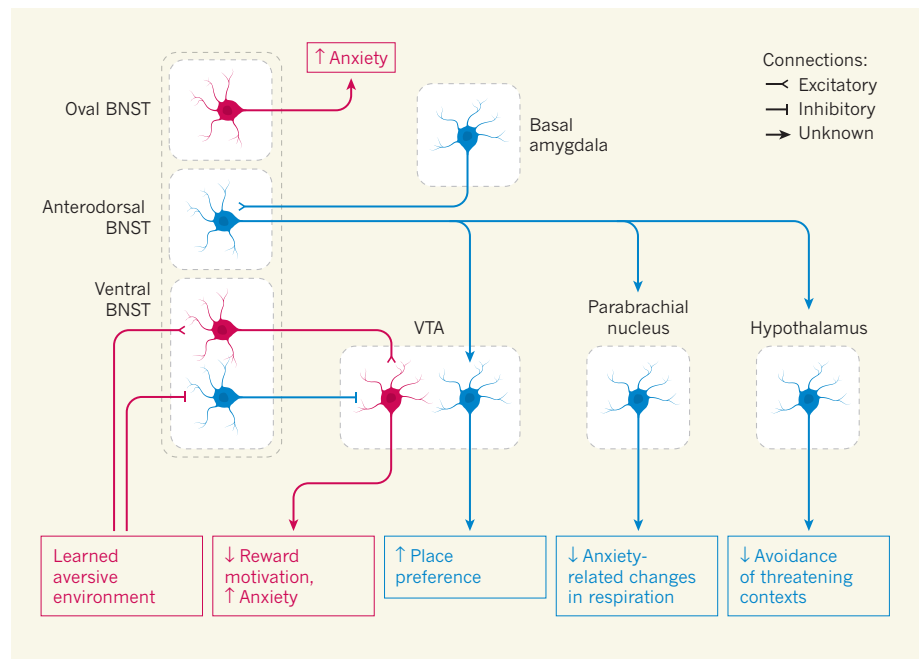


Figure 1 | Multiple personalities of an anxiety circuit. Two studies^{2,3} show that various subregions in the bed nucleus of the stria terminalis (BNST) of the mouse brain contain intermixed cell populations that can produce (red) or ameliorate (blue) anxiety in a modular manner. Within the dorsal BNST, outputs from the oval nucleus promote anxiety, whereas outputs from the anterodorsal BNST — driven by activity in the amygdala — reduce anxiety. Anterodorsal BNST neurons also make specific connections to other brain regions, such as the hypothalamus, parabrachial nucleus and ventral tegmental area (VTA), to ameliorate specific features of anxiety. The ventral BNST contains intermixed but functionally distinct subpopulations of neurons. When the mice are exposed to a known anxiety-causing environment, some of these neurons are excited, stimulating their connections in the VTA to produce anxiety and reduce reward motivation. Other ventral BNST neurons that reduce anxiety and enhance reward motivation are inhibited, facilitating the production of anxiety.

Kim *et al.* also examined specific connections between the adBNST and other brain regions and found that, depending on the connections involved, the adBNST reduced specific aspects of the anxiety response. For instance, stimulating the connections between the adBNST and the hypothalamus reduced the tendency of mice to avoid anxiety-producing places; stimulating connections to neurons of the parabrachial nucleus led to reduced anxiety-induced changes in respiration; and stimulating connections with VTA neurons resulted in place preference (Fig. 1).

The two studies give us a richer understanding of how anxiety is represented by opposing but complementary neural circuits in the BNST. They highlight the modular nature of anxiety circuits and suggest a concerted mechanism for bidirectional regulation of anxiety-related responses. This type of bidirectional coding has been seen in other parts of the anxiety circuit^{10,11}, particularly in the brain’s medial prefrontal cortex, in which single neurons differentially represent safe and anxiety-producing environments.

In fact, this type of circuit design may be a general feature of both fear and anxiety systems. There is strong evidence^{12,13} that partially distinct neuronal subpopulations mediate fear and safety-from-fear learning. Moreover, fear and anxiety are closely related

conceptually, and brain regions such as the amygdala, medial prefrontal cortex, hippocampus and BNST are involved in both. Understanding the principles shared by the two systems, and how their respective neural circuits interact, will be research areas of great interest for the future. ■

Joshua P. Johansen is at the **RIKEN Brain Science Institute, Saitama 351-0198, Japan.**
e-mail: jjohans@brain.riken.jp

1. Kessler, R. C. *et al.* *Arch. Gen. Psychiatry* **62**, 593–602 (2005).
2. Jennings, J. H. *et al.* *Arch. Gen. Psychiatry* **496**, 224–228 (2013).
3. Kim, S.-Y. *et al.* *Nature* **496**, 219–223 (2013).
4. Davis, M., Walker, D. L., Miles, L. & Grillon, C. *Neuropsychopharmacology* **35**, 105–135 (2010).
5. Sullivan, G. M. *et al.* *Neuroscience* **128**, 7–14 (2004).
6. Duvarci, S., Bauer, E. P. & Pare, D. *J. Neurosci.* **29**, 10357–10361 (2009).
7. Poulos, A. M., Ponnusamy, R., Dong, H. W. & Fanselow, M. S. *Proc. Natl Acad. Sci. USA* **107**, 14881–14886 (2010).
8. Kudo, T. *et al.* *J. Neurosci.* **32**, 18035–18046 (2012).
9. Dong, H.-W. & Swanson, L. W. *J. Comp. Neurol.* **468**, 277–298 (2004).
10. Adhikari, A., Topiwala, M. A. & Gordon, J. A. *Neuron* **71**, 898–910 (2011).
11. Tye, K. M. *et al.* *Nature* **471**, 358–362 (2011).
12. Maren, S. & Quirk, G. J. *Nature Rev. Neurosci.* **5**, 844–852 (2004).
13. Herry, C. *et al.* *Eur. J. Neurosci.* **31**, 599–612 (2010).

Intraterrestrial lifestyles

Genome sequencing of cells plucked from marine sediments reveals metabolic details for two abundant lineages of Archaea. These microorganisms may play a key part in breaking down protein buried deep inside Earth. [SEE LETTER P.215](#)

DAVID L. VALENTINE

Vast communities of single-celled organisms reside in a microbial purgatory deep below Earth's surface. They are known as the intraterrestrials. Bounded from below by the inhospitable temperature of Earth's interior, intraterrestrials face a chronic limitation of food-derived energy¹ because they are far removed from sunlight-driven productivity. Once thought barren of life, their subsurface realm is now recognized as a varied habitat influenced as much by geological processes as by the biosphere (Fig. 1). All three domains of life — Bacteria, Eukarya and Archaea — are represented there, as are the viruses that infect them. On page 215 of this issue, Lloyd *et al.*² present genome sequences of four intraterrestrial archaea. The sequences shed light on the history of this enigmatic domain, and suggest that the groups that these organisms represent contribute to the degradation of proteins in sediments blanketing the ocean floor*.

To arrive at these findings, Lloyd and colleagues harvested sediment from the upper fringe of the subsurface biosphere, just a few metres beneath the sea floor in a Danish bay. They separated individual cells from the sediment before sequencing the cells' DNA and reconstructing their genomes by aligning millions of overlapping DNA fragments. This provided incomplete reconstructions of genomes for four archaeal organisms that represent two evolutionary lineages (three of the four are near relatives and one is a distant cousin). The estimated capture of genomic content, ranging over 32–70%, was sufficient to identify each organism's evolutionary placement and core metabolisms but not to rigorously exclude any specific genetic capability. The authors analysed the genomes to identify specific genes and their functions, reconstruct metabolic networks and ascertain evolutionary

*This article and the paper² under discussion were published online on 27 March 2013.

relationships with other organisms.

Our current ability to pluck an individual cell from the environment and sequence most of its genome is a far cry from the manual 'fingerprinting' with sheets of cellulose acetate paper used by the late biologist Carl Woese in the 1970s to first recognize the archaeal domain. Deciphering the evolutionary history of microbes in the face of such technological advances requires iterative refinement. Some bacterial lineages evolve so rapidly that attempting to resolve their deep evolutionary history may prove futile³, and the loss of superfluous genes in organisms that live in

symbiosis with others, such as some hyperthermophilic archaea⁴, may obscure traditional metrics of relatedness. Nonetheless, the evolutionary distinction between Bacteria and Archaea remains firm⁵. And because the organisms sequenced by Lloyd and colleagues have evolved a lifestyle suited to the subsurface environment and paced to change on geological timescales, they should provide a window on the early stages of microbial evolution. Indeed, the genomes reveal ancient evolutionary splits, perhaps to the phylum level for one lineage, and our understanding of these relationships is sure to develop further. We can rely on one certainty, though: the name tags presently assigned to these groups — the miscellaneous crenarchaeotal group and the marine benthic group-D — will soon be replaced with something of Latin etymology.

The authors' comparative analysis of these four genomes points to a common metabolic strategy, which in turn suggests a potentially important role for these organisms and their relatives in the biosphere at large. The core metabolism for both lineages seems to involve the breakdown of proteins for energy generation. Apart from some biochemical differences, it seems that all four archaea do this by breaking apart proteins outside the cell, transporting the resulting peptides into the cell, metabolizing them and conserving the derived energy by pumping cations out of the cell. The portion of metabolism that occurs outside the cell may result from an adaptation to confinement between grains of sediment; using extracellular enzymes may provide access to proteinaceous 'geomolecules' that are too large to enter the cell, with the added bonus that the enzymes remain in the local environment. The diet of one lineage seems to include D-amino acids, which are enriched in the cell walls and spores of some bacteria and so are deposited to and synthesized within the subsurface biosphere⁶. Thus, it seems that members of these lineages decompose protein deposited during sedimentation or formed in the subsurface, including that released by the decay or dismemberment of bacteria and their spores.

The genomes also reveal a dearth of genes that might couple their metabolism to respiration. This intriguing finding is consistent with the archaea using a fermentative metabolism. However, the paucity of protein substrates in the subsurface environment may inhibit the classic Stickland-type amino-acid fermentation⁷ and force these intraterrestrials into syntrophic, or 'cross-feeding', relationships with other organisms amenable to accepting their waste products⁸. Although some of the metabolic details remain



Figure 1 | Microbial purgatory. Jacopo Tintoretto's *Purgatory* depicts a temporary habitat sandwiched between a realm of light above and the inferno of Hell below. But although Earth's subsurface might be deprived of light and experience extreme heat, it is not barren, and the organisms that live there play important parts in our planet's biogeochemistry. The genome sequences of four subsurface archaea, described by Lloyd *et al.*², suggest that these microorganisms contribute to the degradation of proteins in ocean sediments.

ALINARI ARCHIVES/CORBIS

unclear, the high-protein diet proposed for these intraterrestrials may be of broad importance, as it implies that their metabolism may release bound nitrogen that is otherwise recalcitrant in marine sediments.

With a globally distributed population that could be of the order of 10^{27} individuals (based on a relative abundance of these organisms of approximately 0.4% in a total cellular abundance in the sub-sea-floor sediment⁹ of 2.9×10^{29}), it is hard to argue with the evolutionary success of these intraterrestrial archaeal lineages or to marginalize them as relicts of bacterial evolution gone awry. Although circumstantial, the evidence presented by Lloyd *et al.* suggests that these lineages persist despite chronic calorie restriction. Such a lifestyle is consistent with the theory that adaptations to manage chronic energy stress are hallmarks of the ecology

and evolution of the archaeal domain¹⁰.

Lloyd and colleagues' results are exciting not only because they raise questions about the ecology, evolution, metabolism and biogeochemistry of intraterrestrial life, but also because they point to directions that might provide answers. For example, the archaeal genomes hold clues that will guide strategies for cultivating these organisms, and isolation in pure culture would enable studies of their evolution, biochemistry and physiology. Similarly, knowing the genomic content of these archaea will allow environmental surveys and experiments to probe their genetic variability, the abundance and content of their gene transcripts, their protein machinery, their natural levels of metabolic activity and their population dynamics. We are clearly just scratching the surface of the subsurface. ■

STRUCTURAL BIOLOGY

A solution to the telomerase puzzle

Telomerase synthesizes DNA sequences that protect the integrity of chromosome ends. A model for how the components of this enzyme complex co-assemble offers insight into its structure and function. [SEE ARTICLE P.187](#)

BENJAMIN M. AKIYAMA & MICHAEL D. STONE

Early research on the mechanism of DNA replication uncovered a startling fact: each round of cell division slightly shortens the ends of chromosomal DNA, which if left unchecked would eventually lead to the loss of essential genetic material¹. This 'end replication problem' is solved by telomeres — structures at the ends of chromosomes that contain a series of non-coding DNA repeats, and which become shorter themselves but protect the coding regions from damage². To synthesize and maintain telomeres in rapidly dividing cells, the enzyme telomerase comes into play³. The discovery that telomerase has a key role in the immortalization of cancer cells by preventing telomere loss during cancer-cell division⁴ has sparked efforts to understand the chemical structure of this enzyme complex. But despite more than two decades of research, many details of telomerase structure have remained unclear, until Jiang and colleagues' report⁵ on page 187 of this issue*.

Telomerase has been particularly refractory to structural studies because of its naturally low abundance and its complexity. This holoenzyme — a multi-subunit complex consisting

of several proteins and an evolutionarily conserved telomerase RNA — has proven challenging to reconstitute in large quantities *in vitro*. Instead, structural biologists have focused on fragments of the complex that were more easily tamed, providing glimpses into parts of it^{6,7}.

Much of the pioneering structural work was performed on telomerase from the model organism *Tetrahymena thermophila*. More recently, biochemical studies of naturally assembled telomerase complexes from this well-characterized protozoan established the entire complement of molecules that comprise the functional holoenzyme⁷. With the molecular players identified, and partial high-resolution structures accumulating, a void emerged in understanding how the pieces of the telomerase complex assemble. Jiang *et al.* present a structure of the *T. thermophila* telomerase obtained by electron microscopy (EM), providing the first view of how the telomerase puzzle fits together.

The initial structure that the authors solved using EM reconstruction techniques provided a three-dimensional 'envelope' of the holoenzyme's overall structure at 25 ångströms resolution. To identify the position of each subunit within this structure, the researchers turned to an elegant combination of genetic manipulation and biochemical purification

David L. Valentine is in the Department of Earth Science and the Marine Science Institute, University of California, Santa Barbara, Santa Barbara, California 93106, USA. e-mail: valentine@geol.ucsb.edu

1. Hoehler, T. M. & Jørgensen, B. B. *Nature Rev. Microbiol.* **11**, 83–94 (2013).
2. Lloyd, K. G. *et al.* **496**, 215–218 (2013).
3. Shapiro, J. B. *et al. Science* **336**, 48–51 (2012).
4. Huber, H. *et al. Nature* **417**, 63–67 (2002).
5. Koonin, E. V., Puigbò, P. & Wolf, Y. I. *J. Comput. Biol.* **18**, 917–924 (2011).
6. Lomstein, B. A. *et al. Nature* **484**, 101–104 (2012).
7. Stickland, L. H. *Biochem. J.* **28**, 1746–1759 (1934).
8. Adams, C. J., Redmond, M. C. & Valentine, D. L. *Appl. Environ. Microbiol.* **72**, 1079–1085 (2006).
9. Kallmeyer, J. *et al. Proc. Natl Acad. Sci. USA* **109**, 16213–16216 (2012).
10. Valentine, D. L. *Nature Rev. Microbiol.* **5**, 316–323 (2007).

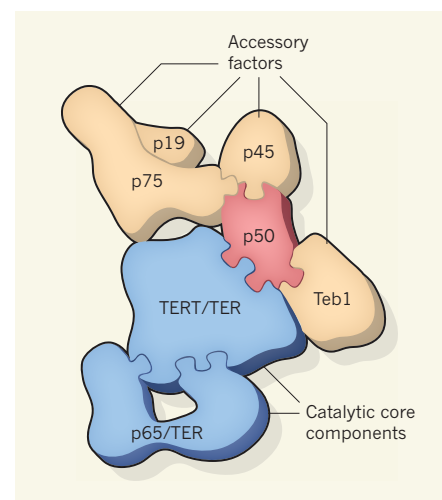


Figure 1 | The telomerase jigsaw puzzle. Jiang and colleagues' electron microscopy data⁵ show that the telomerase holoenzyme is spatially organized into two parts, with a catalytic core on one side and a set of accessory factors on the other. The p50 protein of this enzyme complex serves as a crucial bridge between the two parts. (Figure adapted from ref. 5.)

techniques. This approach allowed site-specific attachment of bulky antibodies to the complex, which could in turn be visualized as additional density in the EM reconstructions when compared with the unlabelled enzyme. Similarly, they engineered a unique RNA sequence into the complex to localize the telomerase RNA. This sequence binds to a viral coat protein that, like the antibody labelling, introduced unambiguous EM density. These experiments were sufficiently specific to identify the locations of all but one of the seven proteins in the complex, as well as a crucial region of the telomerase RNA.

Previous biochemical work suggested that, functionally, telomerase can be divided into two parts: a catalytic core consisting of components essential for the enzyme's DNA-extension

*This article and the paper⁵ under discussion were published online on 3 April 2013.

activity *in vivo*, and a set of accessory factors that promote multiple rounds of telomere-DNA synthesis⁸. Jiang and co-workers' structure reveals that the holoenzyme is spatially organized around this division, with the catalytic core (consisting of the essential telomerase RNA TER and the proteins p65 and TERT) on one side and the accessory factors (p75, p19, p45 and Teb1) on the other (Fig. 1). The authors propose a complete model of the catalytic core that is consistent with existing biochemical data of known inter-subunit interactions, thus providing a validation of their approach.

The structure also reveals a surprise: the previously uncharacterized holoenzyme protein p50 serves as a crucial bridge between the catalytic core and the accessory factors, and, on the basis of *in vitro* studies, has a central role in promoting high levels of telomerase activity. The unexpected importance of the p50 protein in supporting telomerase holoenzyme assembly and function suggests that it should be the subject of future study. In particular, it would be of great interest to identify p50-related proteins in other biological systems, such as vertebrates and yeast, and to determine whether these proteins perform similar roles in mediating the assembly of multi-subunit biological complexes.

The structure reported in the present paper provides a much-anticipated model for telomerase structure and establishes a crucial framework for future structure–function analyses of this holoenzyme. For instance, it proposes sites of RNA–protein and protein–protein interaction that should be characterized. Also, it remains unclear whether the structural organization between DNA-handling and core telomerase factors is preserved in the mammalian telomerase enzyme. A recently solved⁹ EM structure of human telomerase suggests that, unlike *T. thermophila* telomerase, the human enzyme can form a functional dimer. However, the structure gives no information about the interaction of human telomerase with known telomerase-associated DNA-binding proteins.

Ultimately, a complete mechanistic understanding of telomerase assembly and function will require high-resolution structures that provide atomic-level insight into the functional contributions of proteins, RNA and the DNA substrates within the telomerase complex. Given the close link between telomerase activity and cancer, mechanistic information derived from telomerase structural studies could be useful for the design and optimization of anticancer drugs targeting this holoenzyme. Historically, EM reconstructions like that reported by Jiang *et al.* have been milestones on the path towards solving the X-ray crystal structures of other essential cellular machines, including the ribosome and RNA polymerase. So an even clearer picture of the telomerase holoenzyme might be in sight. ■

Benjamin M. Akiyama is in the Department of Molecular, Cell and Developmental Biology

and the Center for Molecular Biology of RNA, University of California, Santa Cruz. **Michael D. Stone** is in the Department of Chemistry and Biochemistry and the Center for Molecular Biology of RNA, University of California, Santa Cruz, Santa Cruz, California 95064, USA. e-mail: mds@ucsc.edu

1. Olovnikov, A. M. *J. Theor. Biol.* **41**, 181–190 (1973).
2. Harley, C. B., Fitcher, A. B. & Greider, C. W.

Nature **345**, 458–460 (1990).

3. Greider, C. W. & Blackburn, E. H. *Cell* **43**, 405–413 (1985).
4. Kim, N. W. *et al. Science* **266**, 2011–2015 (1994).
5. Jiang, J. *et al. Nature* **496**, 187–192 (2013).
6. Mason, M., Schuller, A. & Skordalakes, E. *Curr. Opin. Struct. Biol.* **21**, 92–100 (2011).
7. Theimer, C. A. & Feigon, J. *Curr. Opin. Struct. Biol.* **16**, 307–318 (2006).
8. Min, B. & Collins, K. *Mol. Cell* **36**, 609–619 (2009).
9. Sauerwald, A. *et al. Nature Struct. Mol. Biol.* **20**, 454–460 (2013).

SOLAR SYSTEM

Saturn's ring rain

Saturn's atmosphere bears a latent image of its icy rings, implying that electrically charged bits of water ice are being transported along magnetic-field lines of force from sources in the ring plane to the upper atmosphere. [SEE LETTER P.193](#)

JACK CONNERNEY

The origin and evolution of Saturn's rings has been a particularly vexing problem over the past several decades of planetary exploration, and is one that has obvious implications for the formation of the Solar System. And although much progress has been made in understanding the dynamics of the ring system, both observationally and theoretically, the big questions remain unanswered. How and when did Saturn's rings form? Can they be the last surviving remnants of a much more massive ring system that formed 4.5 billion years ago¹, coeval with the primordial gaseous disk that condensed to form Saturn? Or are the rings relatively young, say about

100 million years old, as implied by a suite of mechanisms that would disperse², darken³ and erode^{4–6} them? Observations obtained by O'Donoghue and colleagues⁷ with the Keck telescope, and reported on page 193 of this issue, point ominously to an electromagnetic erosion mechanism that siphons away ring mass and deposits it in Saturn's upper atmosphere — a process that may explain some of the gross structure of the rings^{4–6}.

Saturn's rings are comprised of nearly pure water-ice objects of all sizes, from submicrometre-sized grains to embedded moonlets kilometres across. However, the vast majority of ring mass, equivalent to that of a 500-km icy sphere, is contained in objects ranging in size from centimetres to a few metres⁸. Such objects reside in

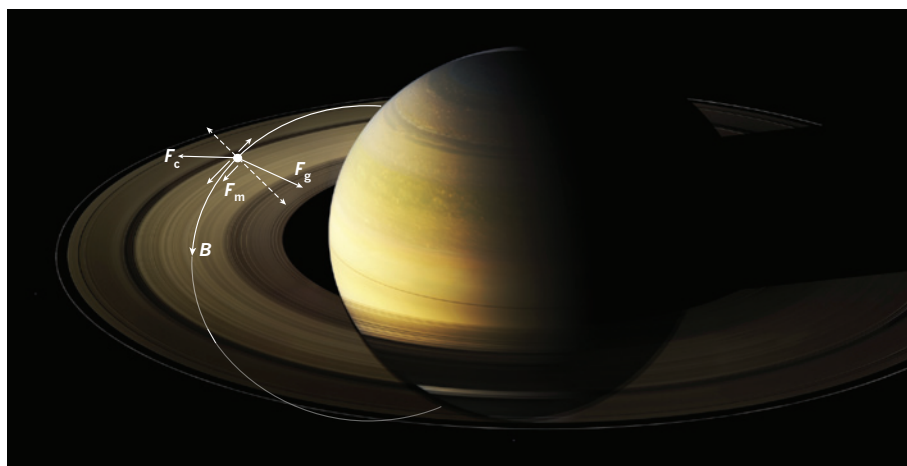


Figure 1 | Saturn's rings and magnetic field. A magnetic field line passing through the 'optically thick' B ring is shown. An electrically charged particle is constrained to move along the magnetic field (B) in response to the components of the gravitational and centrifugal forces (F_g and F_c) that are parallel to the magnetic field, and to the 'magnetic mirror' force (F_m). Dashed vectors represent the components of F_g and F_c in the direction perpendicular to B . Small solid vectors denote the components parallel to B . O'Donoghue and colleagues⁷ observed a reduction in H_3^+ -ion emissions from the region of Saturn's atmosphere that is magnetically linked to the rings, indicating transport of water along the magnetic field lines from the rings to the atmosphere.

CAROL LADD/NASA

Keplerian orbit about Saturn, and their motion lies well within the domain of Newtonian physics (classical mechanics). Indeed, many interesting features in the rings may be explained by the dynamics of a collisional, self-gravitating ensemble of particles confined to a thin disk in orbit about a central body, behaving like a dense gas, characterized by viscosity, temperature and pressure⁹. It may require coupled hydrodynamic and gravitational models, and a fast computer, to describe the collective motion of so many particles, but it remains a problem of classical mechanics.

By contrast, the motion of very small (sub-micrometre-sized) ice particles will be quite different, if they acquire sufficient electrical charge, for example by photoionization or exposure to dense plasma evolving from a micrometeorite impact. Particles with a high charge-to-mass ratio (one electron charge per 1,000 water molecules is sufficient) gyrate about magnetic lines of force in response to the magnetic Lorentz force, which acts in a direction perpendicular to the magnetic field. The motion of such a particle can be described as the combination of a circular motion about the magnetic line of force and the motion of this 'guiding centre' along the magnetic field. In essence, the particle is constrained to move along the magnetic field like a bead on a wire (Fig. 1). These particles will slide along the magnetic field in response to the components of the gravitational and centrifugal forces that are parallel to the magnetic field, and in response to a third force, the 'magnetic mirror' force, which is parallel to the magnetic field and points in the direction of weaker magnetic-field strength (towards the 'magnetic equator'). The latter force is a simple function of the particles' velocity with respect to the magnetic field.

In the case of any other magnetized planet, such forces would quickly disperse small ring particles that acquire a charge. But Saturn is unique among all the magnetized planets of the Solar System in that its magnetic field is symmetric about its rotation axis^{10,11}; there is a unique pair of conjugate latitudes, north and south, that map to a specific radial distance in the ring plane. Mass excavated from the rings in the form of particles of high charge-to-mass ratio, if not returned to the rings and reabsorbed, must therefore be deposited at specific latitudes in Saturn's atmosphere¹² (ring-plane conjugates). The current rate of mass erosion as a function of radial distance in the ring plane could be read¹³ from the variation with latitude of water influx at the top of Saturn's atmosphere — if only one could measure it.

O'Donoghue and colleagues did not measure water influx, but they did observe a good proxy for it: emissions of the H_3^+ ion. Water introduced into the upper atmosphere facilitates the rapid chemical recombination of the major ionospheric (upper atmosphere) ions by charge exchange¹⁴, so a greater depletion of H_3^+ -ion

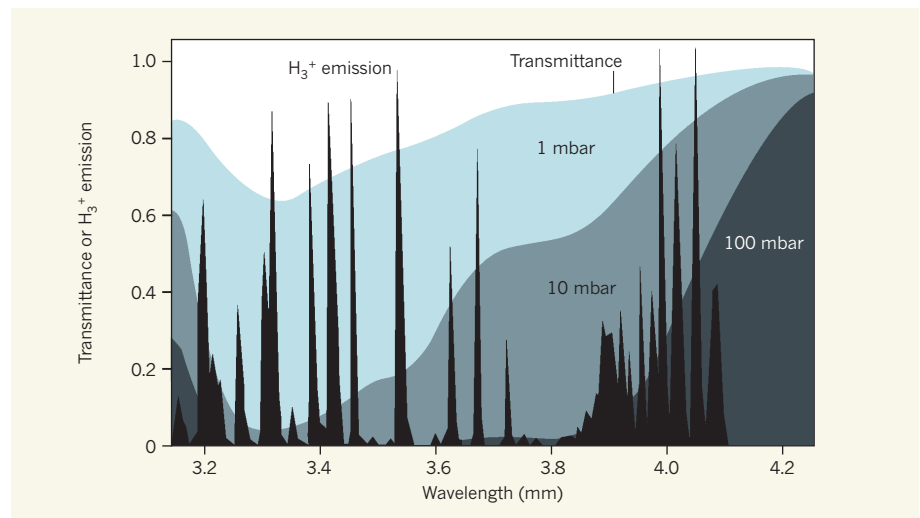


Figure 2 | Saturn's atmospheric transmission and H_3^+ emission lines in the near infrared.

Transmittance is shown for atmospheric pressure levels of 1, 10 and 100 millibar, corresponding to progressively greater depth in Saturn's atmosphere. In this region of the spectrum, light from below (including reflected sunlight) is greatly attenuated by absorption due to methane, so the planet generally looks dark. H_3^+ emission, originating well above the level at which methane is found, escapes without attenuation. H_3^+ emission lines in this region of the spectrum can be viewed against a dark planetary background, making the near infrared an ideal hunting ground for characterizing H_3^+ emissions.

density will be observed at latitudes that receive more water. The authors' measurements clearly show that water is being supplied, along magnetic field lines, to the ionosphere, from sources throughout the ring plane — a ring rain, as it is called. Gaps in the rings are evidently weak sources, not surprisingly, as there is little ring material therein to be eroded. Their measurements also demonstrate that the current ring-erosion rate (as a function of radial distance) differs from that thought to have shaped the C-ring/B-ring boundary⁴ and the inner B-ring transparency^{5,6} over tens of millions of years of evolution. Whether the water is transported in the form of ions or, more efficiently, in the form of charged submicrometre grains is not yet clear. And much work remains to be done before the mass-erosion rate can be worked out, quantitatively, from H_3^+ emission intensities, because it has been difficult to precisely match variations in the observed ionospheric electron density^{15,16} with models that use an exogenous water influx^{17,18}.

The potential of this observational technique for improving our understanding of the electromagnetic erosion of Saturn's rings is particularly exciting. The near-infrared region of the electromagnetic spectrum is replete with many discrete H_3^+ emission lines that span a rather broad methane absorption band (Fig. 2). By judicious choice of emission lines, it should be possible to image ionospheric H_3^+ with high signal-to-noise against a planetary disk darkened by methane absorption deeper in the atmosphere. All that is needed is a telescope with a large effective aperture and time to integrate.

Saturn's rings as observed today probably bear little resemblance to the rings that originally formed. They are highly evolved, much

like our Solar System, albeit over a shorter span of time^{2,4,14}. To understand when and how they formed, one needs to understand the processes that shaped the rings we see today. One of those processes — electromagnetic erosion — has projected an image of the rings upon the disk of Saturn⁷ and perhaps left clues in the rings as well⁴⁻⁶. If so, these brilliant rings have yet a tale to tell, in the language of both Newton and Lorentz. ■

Jack Connerney is at the Planetary Magnetospheres Laboratory, Goddard Space Flight Center, National Aeronautics and Space Administration, Code 695, Greenbelt, Maryland 20771, USA.
e-mail: jack.connerney@nasa.gov

- Canup, R. M. *Nature* **468**, 943–946 (2010).
- Salmon, J., Charnoz, S., Crida, A. & Brahic, A. *Icarus* **209**, 771–785 (2010).
- Cuzzi, J. N. & Estrada, P. R. *Icarus* **132**, 1–35 (1998).
- Northrop, T. G. & Connerney, J. E. P. *Icarus* **70**, 124–137 (1987).
- Ip, W.-H. *J. Geophys. Res.* **88**, 819–822 (1983).
- Northrop, T. G. & Hill, J. R. *J. Geophys. Res.* **88**, 6102–6108 (1983).
- O'Donoghue, J. *et al. Nature* **496**, 193–195 (2013).
- Tiscareno, M. S. *et al. Nature* **440**, 648–650 (2006).
- Cuzzi, J. N. *et al. Science* **327**, 1470–1475 (2010).
- Connerney, J. E. P., Ness, N. F. & Acuña, M. H. *Nature* **298**, 44–46 (1982).
- Burton, M. E., Dougherty, M. K. & Russell, C. T. *Geophys. Res. Lett.* **37**, 24105 (2010).
- Connerney, J. E. P. *Geophys. Res. Lett.* **13**, 773–776 (1986).
- Prangé, R., Fouchet, T., Courtin, R., Connerney, J. E. P. & McConnell, J. C. *Icarus* **180**, 379–392 (2006).
- Connerney, J. E. P. & Waite, J. H. *Nature* **312**, 136–138 (1984).
- Kaiser, M. L., Desch, M. D. & Connerney, J. E. P. *J. Geophys. Res.* **89**, 2371–2376 (1984).
- Fischer, G., Gurnett, D. A., Zarka, P., Moore, L. & Dyudina, U. A. *J. Geophys. Res.* **116**, A04315 (2011).
- Moore, L. *et al. Geophys. Res. Lett.* **33**, L22202 (2006).
- Moore, L. *et al. J. Geophys. Res.* **115**, A11317 (2010).

Glucose–TOR signalling reprograms the transcriptome and activates meristems

Yan Xiong¹, Matthew McCormack¹, Lei Li¹, Qi Hall¹, Chengbin Xiang² & Jen Sheen¹

Meristems encompass stem/progenitor cells that sustain postembryonic growth of all plant organs. How meristems are activated and sustained by nutrient signalling remains enigmatic in photosynthetic plants. Combining chemical manipulations and chemical genetics at the photoautotrophic transition checkpoint, we reveal that shoot photosynthesis-derived glucose drives target-of-rapamycin (TOR) signalling relays through glycolysis and mitochondrial bioenergetics to control root meristem activation, which is decoupled from direct glucose sensing, growth-hormone signalling and stem-cell maintenance. Surprisingly, glucose–TOR signalling dictates transcriptional reprogramming of remarkable gene sets involved in central and secondary metabolism, cell cycle, transcription, signalling, transport and protein folding. Systems, cellular and genetic analyses uncover TOR phosphorylation of E2Fa transcription factor for an unconventional activation of S-phase genes, and glucose–signalling defects in *e2fa* root meristems. Our findings establish pivotal roles of glucose–TOR signalling in unprecedented transcriptional networks wiring central metabolism and biosynthesis for energy and biomass production, and integrating localized stem/progenitor-cell proliferation through inter-organ nutrient coordination to control developmental transition and growth.

Plant photosynthesis, fuelled by solar energy, CO₂ and water to generate renewable organic carbon and oxygen, has a central role in sustaining human life and the ecosystems on Earth. Despite its vital importance, the molecular mechanisms by which photosynthetic products are sensed locally and systematically to activate the metabolic and growth programs in the meristems remain poorly understood^{1,2}. TOR kinase is a master regulator evolutionarily conserved from yeasts to plants and human, that integrates nutrient and energy signalling to promote cell proliferation and growth^{3–5}. Recent research emphasizes the roles of mammalian TOR kinase in translational controls of cell proliferation⁶, insulin signalling^{7,8} and cancer initiation and metastasis⁹. In photosynthetic plants, the molecular functions and the dynamic regulatory mechanisms of TOR kinase remain largely unclear, as the embryo lethality of null *Arabidopsis tor* mutants, partial deficiency of inducible *tor* mutants, and the prevalently perceived rapamycin resistance have hampered genetic and chemical elucidations^{3,4,10}.

At the onset of plant life, the integrated metabolic and developmental programs switch from heterotrophic utilization of maternal seed reserves to photosynthesis-driven metabolic reprogramming and signalling. This switch allows support of potentially infinite plant growth with renewable carbon and energy production in response to CO₂ and sunlight^{11–13}. How plant photosynthetic source and sink organs are coordinated to convey nutrient status, what is the nature of nutrient signals, and how meristems are activated and sustained to continuously supply new cells for growth by photosynthesis are all unknown. To begin to address these fundamental questions, we established a simple and sensitive plant system at the transition checkpoint of heterotrophic to photoautotrophic conversion in *Arabidopsis* seedlings^{13,14}. We applied a combination of chemical, genetics, genomics, bio-computational and cell-based analyses to dissect the TOR signalling networks in meristem activation and plant growth. We discovered that photosynthesis controlled TOR signalling, which was predominantly stimulated by glucose through glycolysis and mitochondrial bioenergetics relays, to rapidly control metabolic transcription

networks and activate the cell cycle in root meristems. Surprisingly, TOR signalling was decoupled from direct glucose sensing via the hexokinase 1 (HXK1) glucose sensor¹¹, growth-hormone signalling¹⁵ and stem-cell maintenance². Our findings establish an unprecedented molecular framework delineating previously unexpected transcriptional regulation of central and secondary metabolic pathways, biogenesis and key regulators of stem and progenitor cell proliferation by TOR kinase. This TOR-regulated molecular framework provides energy, metabolites, biomass, cell cycle machineries and peptide and redox regulators that concertedly drive stem/progenitor-cell proliferation and plant growth through inter-organ nutrient coordination (Supplementary Fig. 1). Integrative systems, cellular and genetic analyses identified E2Fa transcription factor as a novel TOR kinase substrate for an unconventional activation of S-phase genes in cell cycle entry and a determinant of glucose sensitivity in the root meristem. Plant TOR kinase acts as a gatekeeper gauging and linking the photosynthesis-driven glucose nutrient status to comprehensive growth programs through metabolically-regulated signal transduction and transcriptional networks.

Photosynthesis controls root meristems

Although photosynthesis and sugars have a decisive role in root meristem activity^{13,14}, little is known about sugar signalling mechanisms in roots. *Arabidopsis* seedlings germinated in photosynthesis-constrained and sugar-free liquid medium initiated photomorphogenesis, but entered a mitotic quiescent state with arrested root meristem and growth after the depletion of endogenous glucose at 3 days after germination (DAG)^{13,14,16} (Fig. 1a–c). Photosynthesis propelled by higher light and ambient CO₂ was sufficient to promote a rapid and predominantly root growth (Fig. 1a–d). Using the thymidine analogue 5-ethynyl-2'-deoxyuridine (EdU) for *in situ* detection of cell-cycle S-phase entry¹⁷, we showed that the primary root meristem entered mitotic quiescence after the depletion of maternal nutrient supplies, but could be reactivated by photosynthesis (Fig. 1e). Exogenous glucose (15 mM) taken up by root glucose

¹Department of Molecular Biology and Centre for Computational and Integrative Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA. ²School of Life Sciences, University of Science and Technology of China, Hefei 230023, China.

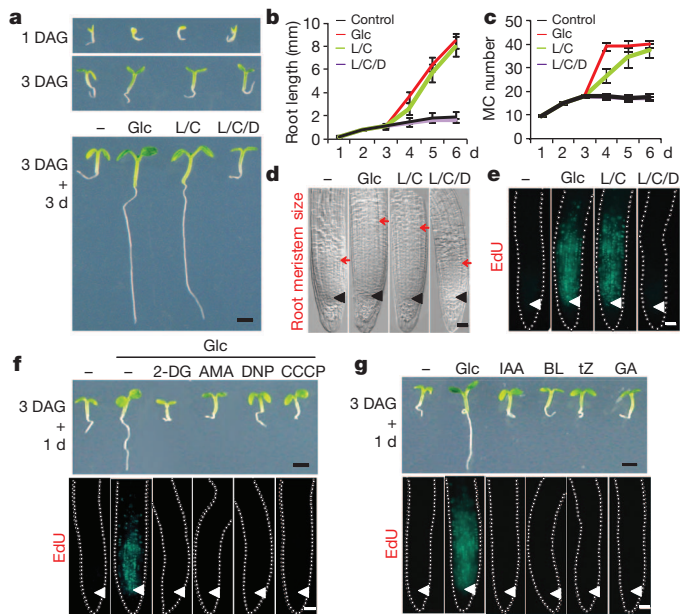


Figure 1 | Photosynthesis controls the metabolic activation of root meristems. **a**, Photosynthesis promotes root growth. DAG, day-after-germination; Glc, glucose; L, light; C, CO₂; D, DCMU. **b–e**, Photosynthesis drives root meristem activation. Results of primary root length, meristem cell (MC) number, meristem size and S-phase entry. Means \pm s.d. ($n \geq 25$). **f**, Glycolysis and mitochondrial bioenergetics stimulate root growth and meristem proliferation. 2-DG, 2-deoxyglucose; AMA, antimycin A; DNP, 2,4-dinitrophenol; CCCP, carbonyl cyanide *m*-chlorophenylhydrazine. **g**, Growth-hormone treatments. IAA, indole-3-acetic acid; BL, brassinosteroid; tZ, trans-zeatin; GA, gibberellins. Scale bar, 1 mm (**a** and **f**, upper panels) or 25 μ m (**d**, **e** and **f**, **g**, bottom panels). Arrowheads, quiescent centre; red arrow, transition zone.

transporters¹⁸ was sufficient to fully substitute for photosynthetic support of root meristem activation and the acceleration of root growth (Fig. 1a–e). The quiescent meristem reactivation occurred within 2 h of glucose treatment (Supplementary Fig. 2). Root growth was quadrupled while the root meristem cell number and size doubled in 24 h (Fig. 1b–g). Consistent with long-distance sucrose and glucose transport through the phloem from shoots to roots^{1,19}, this inter-organ glucose signalling and growth coordination was completely blocked by the herbicide 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) inhibiting photosynthesis (Fig. 1a–e). Only physiological levels of glucose and sucrose, but not other sugars (fructose, xylose and galactose), strongly supported root growth and meristem activation (Supplementary Fig. 3).

Root meristem activation and growth required specific glucose metabolism via glycolysis and mitochondrial electron-transport-chain/oxidative-phosphorylation. The glycolysis blocker 2-deoxyglucose (2-DG), the mitochondrial electron transport inhibitor antimycin A (AMA), and the mitochondria uncouplers, 2,4-dinitrophenol (DNP) and carbonyl cyanide *m*-chlorophenylhydrazine (CCCP)⁵, completely prevented both root growth and root meristem reactivation (Fig. 1f). Reactivation of the quiescent root meristem by light and CO₂ through photosynthesis also relied on the same glycolysis–mitochondria relays (Supplementary Fig. 4), reinforcing the intimate link between photosynthesis and glucose metabolism in supporting meristem activation and sink organ growth.

Plant growth hormones, including auxin, cytokinins, gibberellins and brassinosteroids, have been implicated in controlling the balance between cell division and differentiation, which determines the root meristem size in the presence of sugars^{20,21}. Surprisingly, none of these growth hormones could promote root growth or reactivate the quiescent root meristem at the heterotrophic-to-photoautotrophic transition checkpoint without photosynthesis or exogenous sugars (Fig. 1g). The addition of a physiological mix of amino acids or glutamine also failed to

activate the quiescent root meristem (Supplementary Fig. 5). Our results indicate that glucose acts as the pivotal nutrient signal coordinating leaf photosynthesis and root meristems, and provides a fundamental and evolutionarily conserved metabolic platform through glycolysis–mitochondrial energy relays to supply cellular energetic and signalling requirements for root meristem activation and maintenance.

Glucose–TOR signalling in root meristems

We applied specific chemical inhibitors and chemical genetics to examine the involvement of TOR kinase in root growth and meristem regulation by glucose and photosynthesis¹³. The inducible *tor* mutant had no detectable TOR protein but displayed normal growth during 3 DAG relying on seed nutrients without photosynthesis or exogenous glucose (Supplementary Fig. 6)¹³. However, rapamycin and oestradiol-inducible *tor* mutants blocked the rapid reactivation of the quiescent root meristem in 2 h and the promotion of root growth by glucose at the photoautotrophic transition checkpoint at 3 DAG (Fig. 2a and Supplementary Fig. 2). Based on rapamycin-sensitive phosphorylation of T449 in S6K1 as a conserved indicator of endogenous TOR kinase activity¹³, we revealed that glucose activation of TOR kinase also depended on glycolysis–mitochondria-mediated energy and metabolic relays (Fig. 2b). Significantly, 2-DG, AMA, rapamycin and the *tor* mutant similarly inhibited glucose or light/CO₂ promotion of the doubling of root meristem length and cell number, and *de novo* DNA synthesis visualized by EdU *in situ* staining in 24 h (Fig. 2c, d and Supplementary Figs 3, 4, 7a and 8). Importantly, the *Arabidopsis* glucose sensor HXK1 mutant *gin2* did not prevent the glucose-dependent increase of the meristem cell numbers and *de novo* DNA synthesis (Fig. 2a and Supplementary Figs 7b and 9), which is consistent with the uncoupled signalling and catalytic functions of HXK1¹¹.

Specificity of glucose–TOR signalling

To evaluate the physiological and signalling impact of blocking the mitochondrial electron-transport-chain and TOR kinase by AMA and rapamycin, respectively, in the root meristem, we examined growth-hormone signalling and stem-cell maintenance using well-established marker genes and reporters^{2,15}. Surprisingly, mitochondrial energy relays and TOR kinase activity were decoupled from growth-hormone signalling triggered by auxin and cytokinin (Fig. 3a–d and Supplementary Fig. 10). Quantitative real-time reverse transcriptase PCR (qRT-PCR) analysis revealed that different functional classes of

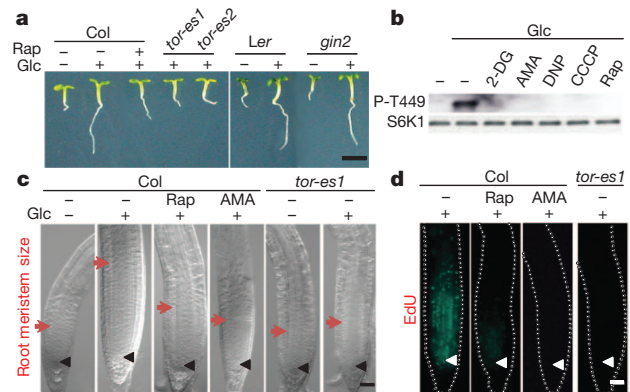


Figure 2 | Glucose–TOR signalling in root meristems. **a**, Root growth activation by glucose is TOR dependent. WT (Col or *Ler*), oestradiol-inducible *tor* mutants (*tor-es1*, *tor-es2*) and *gin2* seedlings at 3 DAG were incubated without or with glucose (Glc) or rapamycin (Rap) for 24 h. Scale bar, 2 mm. **b**, Glucose activates endogenous TOR. TOR activation is detected using an anti-phospho-T449 antibody for S6K1 after 1 h glucose treatment without or with glycolysis inhibitor (2-DG), mitochondrial blockers (AMA, DNP, CCCP) or rapamycin (Rap). **c**, **d**, TOR controls root meristem activation and S-phase entry. Scale bar, 25 μ m.

primary auxin and cytokinin marker genes were all actively induced by hormones in the presence of rapamycin or AMA in wild type, or *tor* seedlings (Fig. 3a, c and Supplementary Fig. 10). Furthermore, rapamycin, AMA or the *tor* mutant did not compromise the activation of auxin and cytokinin signalling green fluorescent protein (GFP) reporters, *DR5::GFP* and *TCS::GFP*, in root meristems (Fig. 3b, d), or perturb the maintenance of the stem-cell niche visualized with the stem-cell and quiescent centre markers, *PLT1::GFP* and *WOX5::GFP*, respectively (Fig. 3e, f). These results unexpectedly indicate that glucose–TOR signalling does not indiscriminately affect general signalling, transcription and translation, but specifically targets cell-cycle regulation of the stem/progenitor cells in the root meristem (Supplementary Fig. 11). Thus, the glucose–TOR signalling network in cell cycle control and root meristem activation is strictly dependent on the glycolysis–mitochondrial energy relays, whereas signalling by plant growth hormones and stem-cell niche maintenance seem to rely on segregated cellular energy and metabolism supporting distinctive transcriptional and translational processes.

Glucose–TOR directs transcriptional networks

To better understand the molecular landscape of the glucose–TOR signalling networks, we performed genome-wide expression profiling to investigate the rapid global transcriptome changes by 2-h glucose (15 mM) treatment in wild-type and *tor* seedlings at the photoautotrophic transition checkpoint. Primary and dynamic glucose response genes were defined by established microarray data analysis algorithms and filtering²², and validated by qRT–PCR analyses of marker genes (Fig. 4, Supplementary Methods and Supplementary Tables 1–5). Based on relatively stringent statistics and filtering (both RMA and dChip, P value < 0.01; signal ratio change $\log_2 \geq 1$, see Supplementary Methods for details), we defined 1,318 up- and 1,050 downregulated *Arabidopsis* genes differentially controlled by a physiological level of glucose (Fig. 4a, c and Supplementary Table 1). The grand scope of reproducible gene expression changes indicated that glucose directs specific and significant transcriptional networks. Strikingly, this swift global transcriptional reprogramming induced by glucose is completely blocked in the inducible *tor* mutant (Fig. 4a, c). It appears crucial to probe the complex and dynamic glucose–TOR-mediated transcriptome in a bioenergetically quiescent checkpoint with minimal background and growth defects before adding specific TOR stimulating signals and inhibitors in multicellular organisms.

Hierarchical clustering analysis of glucose–TOR target genes with ATH1 GeneChip data sets generated by independent research laboratories revealed significant positive correlation of glucose- and sucrose-regulated genes in seedlings^{23,24}. Moreover, adult leaf transcriptome analysis at compensation point [CO_2] (50 p.p.m.) limiting

photosynthesis²⁵ confirmed a negative correlation (Fig. 4a, c). These findings further demonstrated that glucose signal is the main nutrient mediator derived from source leaf photosynthesis for systematic gene regulation and root growth. The sensitivity of our system facilitated the discovery of previously unknown primary glucose target genes, especially enriched in cell cycle and DNA synthesis, transcription, and RNA synthesis/processing among glucose-activated genes (Fig. 4a, b and Supplementary Table 2), and modulating transcription, protein degradation and signalling among glucose-repressed genes (Fig. 4c, d and Supplementary Table 2).

Remarkably, the primary glucose–TOR target genes stratify into a myriad of regulatory and metabolic functional categories, including genes that specifically promote proliferation of stem/progenitor cells for root meristem activation and growth (Fig. 4b, d and Supplementary Tables 1, 3, 4, 6 and 7). Glucose–TOR signalling activated genes encoding root growth factor (RGF) peptides²⁶ (Supplementary Fig. 12) and promoting S-assimilation and glutathione synthesis²⁷ (Supplementary Figs 12 and 14), all essential for cell proliferation in the root meristem. *UPB1* (UPBEAT1) transcription factor, whose overexpression inhibits root meristem expansion through redox control, was repressed²⁰ (Supplementary Fig. 12). Significantly, genes (105) important for cell cycle and DNA synthesis (Mapman²⁸) are highly activated (Fig. 4b and Supplementary Tables 1, 3 and 6).

Over 100 *Arabidopsis* genes encoding ribosomal proteins and protein synthesis machineries were activated by glucose–TOR signalling (Fig. 4b and Supplementary Tables 1 and 3), supporting a universal TOR function in controlling translational processes^{4,5,10,29}. Genes encoding the entire *Arabidopsis* glycolysis and the tricarboxylic acid (TCA) cycle, mitochondrial activities and the electron transport chain were activated by glucose–TOR signalling (Fig. 4b and Supplementary Fig. 13), suggesting a positive feedback loop in TOR-mediated transcriptional control of central carbon and energy metabolism, which is partially conserved in plants, yeasts and mammals^{5,29–31}. TOR kinase also activated genes involved in other key and evolutionarily conserved anabolic processes, including amino acid, lipid and nucleotide synthesis and the oxidative pentose phosphate pathway, that are essential for rapid growth (Fig. 4b and Supplementary Tables 1 and 3), but repressed genes mediating the degradation of proteins, amino acids, lipids and xenobiotic, and autophagy regulation³² (Fig. 4d and Supplementary Tables 1 and 4).

Unique to plant glucose–TOR signalling was its pivotal roles in repressing the metabolic genes for enzymes involved in β -oxidation and glyoxylate cycle required in the germination program of *Arabidopsis* seeds¹², and suppressing catabolic programs for plant survival in the prolonged darkness³² (Fig. 4d and Supplementary Tables 1 and 4). Glucose–TOR signalling also activated broad gene sets coding for the synthesis and modification of plant cell walls, cell-wall proteins (arabinogalactan proteins and expansins), lignin, pectin, secondary metabolites, and a large variety of small molecules²⁸ (Fig. 4b and Supplementary Tables 1 and 3). Notably, key *MYB28/34* transcription factors for the synthesis of glutathione and the indolic/benzoic and aliphatic glucosinolate synthesis pathways (Supplementary Fig. 14), and the genes for lignin and flavonoid synthesis pathways (Supplementary Fig. 15) were also activated by glucose–TOR signalling. Coupled with the extensive TOR regulation of carbon metabolism and biosynthesis was the TOR activation of a large set of genes for protein folding (heat-shock proteins, chaperones and prefoldins), nutrient/metabolite transporters (nitrate transporter and glucose-6-phosphate translocator), lipid transfer proteins, protein secretion and targeting, and vesicle trafficking, but downregulation of genes for various sugar transporters (STP and SWEET)^{1,18} (Supplementary Tables 1, 3 and 4). The plant glucose–TOR signalling networks also integrated a large number of transcription factors, chromatin modulators, signalling regulators and growth- and stress-related proteins that could be unique to plants or conserved in eukaryotes (Fig. 4b, d and Supplementary Table 1). Our findings

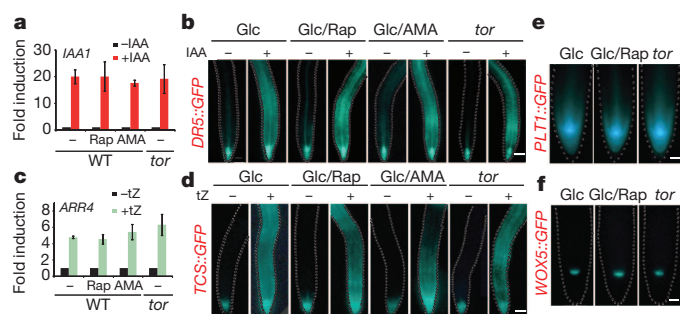


Figure 3 | Auxin and cytokinin signalling and root stem-cell maintenance are decoupled from TOR activation. a, b, Auxin signalling. c, d, Cytokinin signalling. Rapamycin (Rap), mitochondrial blocker (AMA). Primary auxin and cytokinin marker genes were activated by 1 h of indole-3-acetic acid (IAA) or trans-zeatin (tZ) treatment, and analysed by qRT–PCR. Means \pm s.d., $n = 3$. *DR5::GFP* or *TCS::GFP* was activated by 6 h of IAA or tZ treatment. Scale bar, 50 μm . e, f, Root stem-cell maintenance is TOR independent. *PLT1::GFP*, root stem-cell marker; *WOX5::GFP*, root quiescent centre marker. Scale bar, 20 μm .

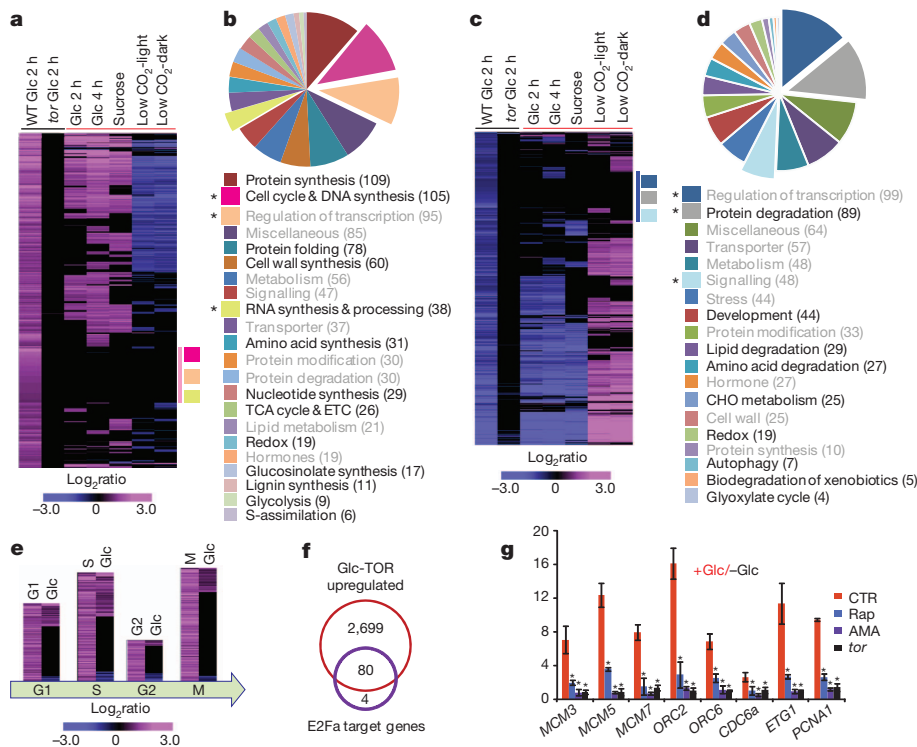


Figure 4 | Glucose–TOR signalling orchestrates transcriptome reprogramming. **a, b**, Glucose–TOR-activated genes. **c, d**, Glucose–TOR-repressed genes. 3 DAG WT or *tor* seedlings were treated without or with glucose (Glc) for 2 h.

Hierarchical clustering analysis of glucose–TOR genes and five independent data sets (Glc 2 h, Glc 4 h, sucrose, low CO_2 -light, low CO_2 -dark). Deep-pink/blue bar indicates novel glucose–TOR genes. The enriched functional categories are highlighted in bold (Supplementary Tables 1, 3 and 4). ETC, electron transport chain. CHO, carbohydrate. **e**, Hierarchical clustering analysis of glucose–TOR genes (Glc) and cell cycle genes (G1, S, G2, M)³³. **f**, Glucose–TOR-activated genes overlap with E2Fa target genes. **g**, Glucose–TOR activates S-phase genes. qRT–PCR analyses. Means \pm s.d., $n = 3$. $^*P < 0.05$.

uncover a previously unanticipated central role of TOR in glucose and energy signalling through rapid transcriptome reprogramming, which is beyond the conventional emphasis on translational controls for mammalian TOR kinase via 4E-BP and S6K^{5,6,9}.

Novel glucose–TOR–E2Fa regulatory relay

To further explore the molecular mechanism by which glucose–TOR signalling controls cell proliferation for meristem activation and root growth, we compared our glucose–TOR target genes with reported cell-cycle oscillation genes³³ using relaxed stringency (both RMA and dChip, P value < 0.01) (Supplementary Table 5). Hierarchical clustering analysis revealed that many glucose–TOR-activated genes matched the typical G1- and S-phase genes (Fig. 4e and Supplementary Table 6). As E2F transcription factors are conserved key regulators of S-phase genes governing cell cycle progression and DNA replication in plants and mammals, we performed stringent computational analyses to identify putative *Arabidopsis* E2Fa target genes, which were defined by E2Fa co-expression (Genevestigator), activation by E2Fa induction in transgenic plants, and possessing putative E2F-binding sites in promoter regions^{34–36} (Supplementary Fig. 16 and Supplementary Table 7). A subset of glucose–TOR-activated genes notably overlapped (95%) with the putative *Arabidopsis* E2Fa target genes (Fig. 4f). Glucose rapidly activated *ORC2/6* (*ORIGIN RECOGNITION COMPLEX*), *MCM3/5/7* (*MINOCHROMOSOME MAINTENANCE*), *CDC6* (*CELL DIVISION CYCLE*), *ETG1* (*E2F TARGET GENE*) and *PCNA1* (*PROLIFERATING CELL NUCLEAR ANTIGEN*), which were significantly diminished in the *tor* mutants or by rapamycin, 2-DG or AMA treatment in wild-type seedlings (Fig. 4g and Supplementary Fig. 18), but not in the glucose sensor *gin2* mutant (Supplementary Fig. 17). Consistently, glucose or sucrose but not other sugars activated these E2Fa target genes (Supplementary Fig. 18), indicating that the dynamic glucose–TOR signalling may partially execute its cell proliferation regulation through E2Fa transcription factor.

E2Fs are the well-established targets of the universal CYC-CDK-RBR (CYCLIN-CYCLIN-DEPENDENT KINASE-RETINOBLASTOMA-RELATED PROTEIN) cascade initiating cell cycle^{33–36}. To explore the novel regulatory link between TOR kinase and E2Fa, we developed a sensitive cell-based assay, in which ectopic expression of

E2Fa alone was sufficient to activate S-phase-specific marker genes in non-dividing and fully differentiated leaf cells (Fig. 5a). S-phase gene activation by E2Fa and T449 phosphorylation in S6K1 were inhibited by rapamycin, AMA or the *tor* mutant (Fig. 5a). Significantly, immunoprecipitated endogenous TOR kinase from *Arabidopsis* plants directly phosphorylated E2Fa *in vitro* (Fig. 5b), which was completely blocked by a specific ATP-competitive TOR kinase inhibitor, torin1³⁷ (Fig. 5b). Consistently, torin1 inhibited T449 phosphorylation of S6K1 *in vivo* and S-phase gene activation by E2Fa in non-dividing leaf cells (Supplementary Fig. 19). This plant TOR kinase also phosphorylated the human 4E-BP1 *in vitro* (Fig. 5b), and thus seemed to resemble the rapamycin-sensitive mammalian-TOR-complex1 (mTORC1) but not mTORC2^{4,6,10}. Based on the differential specificity of PK inhibitors^{7,8,37}, we further demonstrated that E2Fa was a direct substrate of TOR kinase but not the TOR-activated S6K1, which could be inhibited by staurosporine but not torin1 (Fig. 5c). As a direct substrate, E2Fa co-immunoprecipitated with TOR in cells (Fig. 5d). These results indicate that direct E2Fa protein phosphorylation by TOR kinase may be a key step for glucose activation of S-phase genes bypassing or acting downstream the conventional CYC-CDK-RBR cascade.

To map the TOR kinase phosphorylation region(s) in E2Fa, various truncated E2Fa proteins were generated, including the amino-terminal putative regulatory, DNA-binding-dimerization, and the carboxy-terminal transcription activation/RBR-interacting domains^{33–36} (Fig. 6a). The *in vitro* kinase assay revealed that the TOR kinase phosphorylation site(s) are located in the N-terminal 80-residue regulatory domain (Fig. 6a). Surprisingly, removal of the previously defined C-terminal transcription-activation/RBR-interacting domain did not abolish E2Fa activation of S-phase target genes, whereas deletion of the N-terminal TOR kinase phosphorylation region rendered E2Fa inactive without affecting protein translation/stability (Fig. 6a, b). Interestingly, the DNA binding domain alone without TOR phosphorylation was sufficient as the full-length E2Fa for binding to the predicted E2Fa-binding motifs located in the *MCM5* and *ETG1* promoters based on real-time chromatin-immunoprecipitation-qPCR (ChIP-qPCR) analyses (Fig. 6c). These results indicated a novel mechanism of TOR phosphorylation in regulating the activity of E2Fa in transcriptional activation, probably independent of S6K, RBR or translational control (Figs 5 and 6a–c).

The Pro-rich 80 residues contained 16 Ser/Thr residues that could potentially serve as TOR phosphorylation sites^{7,8} (Supplementary Fig. 20). Systematic mutagenesis analyses of the 16 Ser/Thr in eight clusters (Supplementary Fig. 20a) did not reveal dominant TOR kinase phosphorylation sites for the E2Fa activity in target gene activation, suggesting combinatorial or redundant TOR phosphorylation illustrated by mammalian 4E-BP1 and Grb10^{5,7,8}. The mutation of all 16 Ser/Thr residues significantly diminished E2Fa activity (Supplementary Fig. 20b).

To substantiate the genetic link and independently evaluate this surprising glucose–TOR–E2Fa signalling cascade in root meristem activation, we screened and isolated a null allele of the *e2fa* mutant (Supplementary Fig. 21). The truncated E2Fa protein failed to activate target genes (Supplementary Fig. 21c). In the absence of glucose, no overt difference was observed between wild type and *e2fa* in root length and meristem cell number. In contrast, glucose-promoted root growth, root meristem expansion and EdU staining were all significantly compromised in *e2fa* (Fig. 6d and Supplementary Fig. 22a, b). However, other related E2Fs might provide partially overlapping functions^{34,35} (Supplementary Fig. 22c, d), which will require detailed investigations. Several reported *e2fa* RNA interference and insertion mutants independently confirmed similar root meristem and growth defects³⁸. qRT–PCR analysis demonstrated that *e2fa* displayed specifically diminished glucose sensitivity in TOR activation of S-phase genes in the root meristem (Fig. 6e), providing compelling genetic evidence for a key role of E2Fa, together with RGFs and UPB1, in the glucose–TOR transcriptional networks governing root meristem activation (Fig. 6f).

Discussion

Comprehensive chemical, genetic, genomic and systems analyses in *Arabidopsis* seedlings at the photoautotrophic transition checkpoint with minimal TOR signalling background was crucial to lead to our discovery of previously unexpected glucose–TOR transcriptional networks. These networks dynamically repress the transcription programs associated with seed nutrient metabolism for germination, and simultaneously stimulate and sustain the meristem activity for infinite root growth through photosynthesis-driven glucose–TOR signalling. Although TOR was first discovered in yeast, the atypical fermentation lifestyle evolved opposite transcriptional regulation of glycolytic and TCA-cycle genes by yeast TOR1³⁰. Limited evidence supports mammalian TOR signalling in direct transcriptional control³⁹. Current studies emphasize translational regulation by mammalian

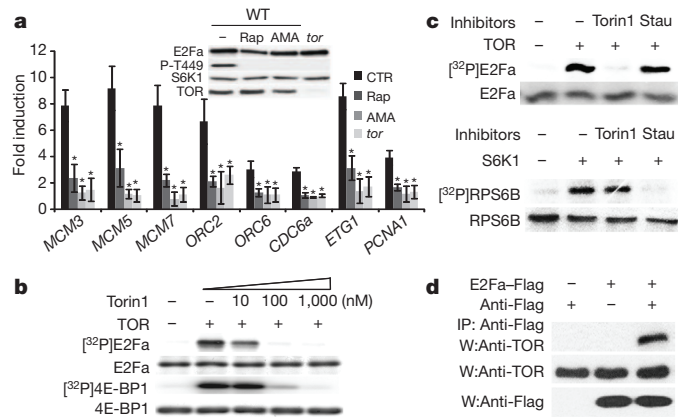


Figure 5 | TOR kinase phosphorylates and activates E2Fa. **a**, Ectopic E2Fa activation of S-phase genes requires glucose–TOR signalling in leaf cells. WT or *tor* protoplasts expressing E2Fa–HA or S6K1–Flag were treated without or with rapamycin (Rap) or antimycin A (AMA). qRT–PCR analyses. P–T449 indicates endogenous TOR kinase activity. Protein blot analysis (inset). Means \pm s.d., $n = 3$. * $P < 0.05$. **b**, **c**, TOR kinase directly phosphorylates E2Fa and 4E-BP1. Torin1 specifically inhibits TOR kinase. Staurosporine (Stau) inhibits S6K1 kinase. **d**, TOR directly interacts with E2Fa by immunoprecipitation (IP) and Western (W) blot analysis.

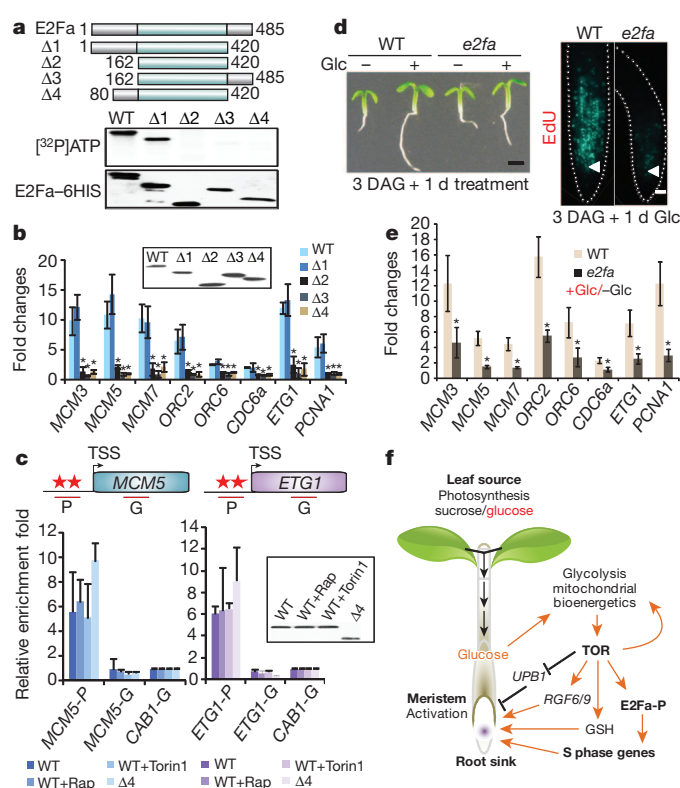


Figure 6 | TOR kinase controls the activity of E2Fa in transcriptional activation. **a**, TOR kinase phosphorylates the N-terminal domain of E2Fa. **b**, TOR kinase phosphorylation is critical for E2Fa activation of S-phase genes. **c**, E2Fa–DNA binding is not affected by TOR kinase phosphorylation. ChIP–qPCR analyses with P (promoter) or G (gene body) primers. Stars, putative E2Fa-binding motifs. Error bars ($n = 2$). **d**, **e**, Glucose responses is diminished in *e2fa* root meristems. Scale bar, 1 mm (left) or 20 μ m (right). qRT–PCR analyses. **f**, Model of leaf–root coordination in glucose–TOR signalling. GSH, GLUTATHIONE; RGF, ROOT GROWTH FACTOR; UPB1, UPBEAT1. Means \pm s.d., $n = 3$. * $P < 0.05$.

TOR via 4E-BP1 and S6K1 phosphorylation, which indirectly modulate limited messenger RNAs and target genes^{5,6,9,31}. Unravelling the plant glucose–TOR signalling networks in metabolic and cell-cycle controls may illuminate the unexplored mTORC1 transcription networks for inter-organ nutrient coordination in animals or in human cancers.

TOR signalling has mainly been linked to amino-acid sensing and insulin/growth regulator signalling to modulate translational controls in mammals^{5,7–9}. As glucose is a universal fuel and metabolic/biomass precursor for most cells, glucose activation of TOR kinase as a central transcriptional regulator of gene sets involved in glycolysis, TCA cycle, ribosome biogenesis, and the synthesis of proteins, amino acids, lipids and nucleotides are probably conserved in multicellular eukaryotes from plants to humans. Our findings establish a molecular framework for future exploration of transcriptional regulators as new TOR kinase substrates coordinating the genes participating in the most conserved and central metabolic pathways in bioenergetics and biosynthesis fundamental to all multicellular organisms. Glucose–TOR signalling also controls plant-specific genes that are uniquely required for plant growth, defence or communication to promote fitness, adaptation and survival. The molecular wiring of the ancient TOR signalling network controls both conserved and divergent metabolic pathways supporting the diverse lifestyle of different organisms.

We discovered E2Fa as a novel TOR kinase substrate, which transcriptionally activates S-phase genes as primary glucose–TOR target genes. The finding breaks the conventional concept of cell-cycle regulation based on the evolutionarily conserved CYC–CDK–RBR–E2F cascade^{34,35}. The direct TOR–E2F link may provide an alternative

entry point of the cell cycle through glucose signalling in the meristem of other plant organs and in other eukaryotes. The successful identification of transcription factors as direct TOR kinase substrates offers an innovative approach for future discovery of unconventional TOR kinase substrates with complex and combinatorial phosphorylation sites. Interestingly, plant growth hormones activate transcription and translation, but are ineffective in mediating cell proliferation in the absence of glucose–TOR signalling (Figs 1g and 3). We propose that glucose–TOR signalling provides essential energy, metabolites, biomass and cell-cycle machineries through concerted transcriptional activation in stem/progenitor cells (Figs 4, 5 and 6). This may explain the prerequisite, fundamentally indispensable and global roles of glucose–TOR signalling in proliferation and growth. Endogenous plant hormones, signalling in specific cells and contexts (Fig. 3)^{2,15}, may modulate specific cell cycle regulators and bring cell-cycle connections to patterning and developmental programs when nutrients and glucose–TOR signalling are available. The findings on glucose–TOR signalling unravel a missing link in nutrient regulation of meristems in plant growth.

METHODS SUMMARY

Quiescent root meristem analyses. To define the transition checkpoint of heterotrophic to photoautotrophic conversion, root meristem size and cell number were carefully monitored daily under microscope for 6 days (Leica, DM5000).

Gene expression analysis. Global gene expression analyses were performed with *Arabidopsis* ATH1 GeneChip arrays (Affymetrix). For details on data processing and analyses, see Supplementary Information.

Full Methods and any associated references are available in the online version of the paper.

Received 27 August 2012; accepted 20 February 2013.

Published online 31 March 2013.

- Chen, L. Q. *et al.* Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* **335**, 207–211 (2012).
- Aichinger, E., Kornet, N., Friedrich, T. & Laux, T. Plant stem cell niches. *Annu. Rev. Plant Biol.* **63**, 615–636 (2012).
- Baena-González, E. & Sheen, J. Convergent energy and stress signaling. *Trends Plant Sci.* **13**, 474–482 (2008).
- Robaglia, C., Thomas, M. & Meyer, C. Sensing nutrient and energy status by SnRK1 and TOR kinases. *Curr. Opin. Plant Biol.* **15**, 301–307 (2012).
- Laplanche, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
- Dowling, R. J. *et al.* mTORC1-mediated cell proliferation, but not cell growth, controlled by the 4E-BPs. *Science* **328**, 1172–1176 (2010).
- Hsu, P. P. *et al.* The TOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* **332**, 1317–1322 (2011).
- Yu, Y. *et al.* Phosphoproteomic analysis identifies Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. *Science* **332**, 1322–1326 (2011).
- Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).
- Moreau, M. *et al.* Mutations in the *Arabidopsis* homolog of LST8/GβL, a partner of the target of rapamycin kinase, impair plant growth, flowering, and metabolic adaptation to long days. *Plant Cell* **24**, 463–481 (2012).
- Moore, B. *et al.* Role of the *Arabidopsis* glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* **300**, 332–336 (2003).
- Graham, I. A. Seed storage oil mobilization. *Annu. Rev. Plant Biol.* **59**, 115–142 (2008).
- Xiong, Y. & Sheen, J. Rapamycin and glucose-target of rapamycin (TOR) protein signaling in plants. *J. Biol. Chem.* **287**, 2836–2842 (2012).
- Sanz, L. *et al.* The *Arabidopsis* D-type cyclin CYCD2;1 and the inhibitor ICK2/KRP2 modulate auxin-induced lateral root formation. *Plant Cell* **23**, 641–660 (2011).
- Müller, B. & Sheen, J. Cytokinin and auxin interaction in root stem-cell specification during early embryogenesis. *Nature* **453**, 1094–1097 (2008).
- Zhang, Z. W. *et al.* The plastid hexokinase pHXK: a node of convergence for sugar and plastid signals in *Arabidopsis*. *FEBS Lett.* **584**, 3573–3579 (2010).
- Kotogány, E., Dudits, D., Horváth, G. V. & Ayaydin, F. A rapid and robust assay for detection of S-phase cell cycle progression in plant cells and tissues by using ethynyl deoxyuridine. *Plant Methods* **6**, 5 (2010).
- Chaudhuri, B. *et al.* Protonophore- and pH-insensitive glucose and sucrose accumulation detected by FRET nanosensors in *Arabidopsis* root tips. *Plant J.* **56**, 948–962 (2008).
- Hirose, N. *et al.* Regulation of cytokinin biosynthesis, compartmentalization and translocation. *J. Exp. Bot.* **59**, 75–83 (2008).
- Tsukagoshi, H., Busch, W. & Benfey, P. N. Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. *Cell* **143**, 606–616 (2010).
- Moubayidin, L. *et al.* The rate of cell differentiation controls the *Arabidopsis* root meristem growth phase. *Curr. Biol.* **20**, 1138–1143 (2010).
- Boudsocq, M. *et al.* Differential innate immune signalling via Ca²⁺ sensor protein kinases. *Nature* **464**, 418–422 (2010).
- Gonzali, S. *et al.* Identification of sugar-modulated genes and evidence for *in vivo* sugar sensing in *Arabidopsis*. *J. Plant Res.* **119**, 115–123 (2006).
- Li, Y. *et al.* Establishing glucose- and ABA-regulated transcription networks in *Arabidopsis* by microarray analysis and promoter classification using a Relevance Vector Machine. *Genome Res.* **16**, 414–427 (2006).
- Bläsing, O. E. *et al.* Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *Plant Cell* **17**, 3257–3281 (2005).
- Matsuzaki, Y., Ogawa-Ohnishi, M., Mori, A. & Matsubayashi, Y. Secreted peptide signals required for maintenance of root stem cell niche in *Arabidopsis*. *Science* **329**, 1065–1067 (2010).
- Vernoux, T. *et al.* The *ROOT MERISTEMLESS1/CADMIUM SENSITIVE2* gene defines a glutathione-dependent pathway involved in initiation and maintenance of cell division during postembryonic root development. *Plant Cell* **12**, 97–110 (2000).
- Thimm, O. *et al.* MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939 (2004).
- Urban, J. *et al.* Sch9 is a major target of TORC1 in *Saccharomyces cerevisiae*. *Mol. Cell* **26**, 663–674 (2007).
- Hardwick, J. S., Kuruvilla, F. G., Tong, J. K., Shamji, A. F. & Schreiber, S. L. Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. *Proc. Natl Acad. Sci. USA* **96**, 14866–14870 (1999).
- Düvel, K. *et al.* Activation of a metabolic gene regulatory network downstream of mTOR complex 1. *Mol. Cell* **39**, 171–183 (2010).
- Baena-González, E., Rolland, F., Thevelein, J. M. & Sheen, J. A central integrator of transcription networks in plant stress and energy signalling. *Nature* **448**, 938–942 (2007).
- Menges, M., Hennig, L., Gruissem, W. & Murray, J. A. Genome-wide gene expression in an *Arabidopsis* cell suspension. *Plant Mol. Biol.* **53**, 423–442 (2003).
- de Jager, S. M. *et al.* Dissecting regulatory pathways of G1/S control in *Arabidopsis*: common and distinct targets of CYCD3;1, E2Fa and E2Fc. *Plant Mol. Biol.* **71**, 345–365 (2009).
- Naouar, N. *et al.* Quantitative RNA expression analysis with Affymetrix Tiling 1.0R arrays identifies new E2F target genes. *Plant J.* **57**, 184–194 (2009).
- Vandepoel, K. *et al.* Genome-wide identification of potential plant E2F target genes. *Plant Physiol.* **139**, 316–328 (2005).
- Liu, Q. *et al.* Kinome-wide selectivity profiling of ATP-competitive mammalian target of rapamycin (mTOR) inhibitors and characterization of their binding kinetics. *J. Biol. Chem.* **287**, 9742–9752 (2012).
- Magyar, Z. *et al.* *Arabidopsis* E2FA stimulates proliferation and endocycle separately through RBR-bound and RBR-free complexes. *EMBO J.* **31**, 1480–1493 (2012).
- Cunningham, J. T. *et al.* mTOR controls mitochondrial oxidative function through a YY1-PGC-1α transcriptional complex. *Nature* **450**, 736–740 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Dai and J. Avruch for S6K antibodies and advice, J. L. Celenza for stimulating discussion, M. D. Curtis and Y. J. Niu for the oestradiol-inducible binary vector, L. Li and J. Bush for seeds and plants, B. Müller for TCS::GFP, J. Friml for DR5::GFP, N. S. Gray and D. M. Sabatini for torin1, and J. F. Li, H. Lee and M. Ramon for critical reading of the manuscript. Y.X. is supported by the MGH Tosteson Postdoctoral Fellowship. C.X. is supported by Chinese Academy of Sciences (KSCX3-YW-N-007). The Research is supported by the NSF, NIH and WJC Special Project (PJ009106) RDA-Korea to J.S.

Author Contributions Y.X. and J.S. initiated the project and designed the experiments; Y.X. carried out most of the experiments; L.L. and Y.X. conducted quantitative ChIP-PCR analyses; Y.X., M.M. and J.S. analysed the microarray data. C.X. isolated the e2fa mutant. Q.H. generated PLT::GFP and WOX5::GFP transgenic lines. Y.X., M.M. and J.S. wrote the manuscript.

Author Information All microarray data are available at the Gene Expression Omnibus under accession number GSE40245. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.X. (xiong@molbio.mgh.harvard.edu) or J.S. (sheen@molbio.mgh.harvard.edu).

METHODS

Plant growth conditions. If not otherwise indicated, all plant materials were grown in a plant growth room with conditions maintained at 23 °C, 65% humidity and 75 $\mu\text{mol m}^{-2}$ s light intensity under 12 h light/12 h dark photoperiod.

Plant materials. Col-0 and *Ler* were used as wild-type *Arabidopsis* plants. *PLT1::GFP*, *WOX5::GFP* and *gin2* are in *Ler* background. All other transgenic plants are in Col-0 background. Oestradiol-inducible *tor* RNAi lines and S6K1-HA overexpression lines were described previously¹³. To generate transgenic *WOX5::GFP* and *PLT1::GFP* lines, the 4.7-kilobase (kb) *WOX5* (*At3g11260*) promoter region and 4.5-kb *PLT* (*At3g20840*) promoter region, respectively, were cloned into an expression vector derived from the pCB302 minibinary vector to drive HXK1–GFP expression. *DR5::GFP*, *TCS::GFP*, *WOX5::GFP* and *PLT1::GFP* lines¹⁵ were crossed with *tor-es1* to generate *DR5::GFP/tor*, *TCS::GFP/tor*, *WOX5::GFP/tor* and *PLT1::GFP/tor* lines. Oestradiol (10 μM) was used to induce TOR depletion, which was confirmed by a specific *Arabidopsis* TOR antibody¹³. The *e2fa* mutant was isolated and confirmed from the *wiscDsLox434F1* line.

Analyses of root meristem reactivation and root growth. *Arabidopsis* seeds (6 seeds per well) were germinated in 6-well plates containing 1 ml of glucose-free liquid medium (0.5 \times MS, pH 5.7 adjusted with KOH) for 3 days to enter the mitotically quiescent state. Quiescent seedlings were treated with glucose (15 mM), plant hormones, amino acid mix (0.1 mM/each) or glutamine (0.1 mM) for the indicated time to reactivate the quiescent root meristem. The concentrations of plant hormones were chosen based on their ability for promoting cell cycle: indole-3-acetic acid (IAA, 0.5 nM)⁴⁰, trans-zeatin (tZ, 100 nM)^{15,40}, gibberellins (GA, 2 μM)⁴¹, and brassinosteroid (BL, 0.01 nM)⁴². Amino acid mix contains 17 amino acids including alanine, arginine, aspartic acid, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tyrosine, valine and cystine. Amino acids failed to activate quiescent root meristem even with high concentration: amino acid mix (1 mM each) or glutamine (0.5 mM and 5 mM, data not shown).

Chemical inhibitor treatments. Quiescent seedlings were pretreated with rapamycin (10 μM), AMA (5 μM), 2-DG (15 mM), DNP (50 μM) or CCCP (10 μM) for 1 h before other treatments. The rapamycin effect is facilitated in the liquid medium¹³.

Enhanced photosynthesis assays. For analysing the effect of photosynthesis on root growth and meristem establishment, quiescent seedlings in glucose-free liquid medium 3 DAG were transferred to glucose-free solid medium (0.5 \times MS and, pH 5.7, 1% agarose) without/with DCMU (20 μM), and grown vertically at 23 °C under constant light conditions of 300 $\mu\text{mol m}^{-2}$ s light intensity for 3 days. To study the molecular link between photosynthesis and glucose–TOR signalling, the quiescent WT or *tor* seedlings were incubated in 400 μl glucose-free liquid medium (0.5 \times MS and 6 mM Na_2CO_3 , pH 5.7) in 6-well plates without/with DCMU (20 μM), 2-DG (15 mM), AMA (5 μM), or rapamycin (10 μM), for 24 h with 200 $\mu\text{mol m}^{-2}$ s light intensity.

Analyses of auxin and cytokinin signalling. For activation of auxin and cytokinin primary marker genes, quiescent seedlings were pretreated without/with rapamycin (10 μM) or AMA (5 μM) for 1 h. Glucose (15 mM) was then added for

2 h, followed by IAA (100 nM) or tZ (100 nM) for an additional 1 h. To analyse *DR5::GFP* and *TCS::GFP* reporter lines, quiescent seedlings were pretreated without/with rapamycin (10 μM) or AMA (5 μM) for 1 h. Glucose (15 mM) was then added for 18 h, followed by IAA (100 nM) or tZ (100 nM) for an additional 6 h (for a total of 24 h incubation). For the inducible *tor* mutant, oestradiol (10 μM) was added at the beginning of germination¹³.

Analyses of stem-cell maintenance. For analysing root stem cells and the quiescent centre, *PLT1::GFP* and *WOX5::GFP* transgenic seedlings were germinated without/with rapamycin for 3 days in 0.5 \times MS medium with 15 mM glucose.

Protoplast transient expression assay. Protoplast transient expression assays were carried out as described^{13,15,22,32}. Data were generated from at least three independent experiments with consistent results. Protoplasts (10⁵) were transfected with 5 μg E2Fa or its truncated variants and incubated for 4 h in 5 ml of mannitol (0.5 M) and KCl (20 mM) buffer (4 mM MES, pH 5.7) in Petri dish (100 mm \times 20 mm) for gene activation analysis. For experiments analysing the effect of rapamycin, AMA and torin1 on E2Fa activated S-phase gene expression, protoplasts were pretreated without/with rapamycin (1 μM), AMA (5 μM) or torin1 (100 nM) for 1 h before E2Fa transfection.

EdU (5-ethynyl-2'-deoxyuridine) staining. EdU staining was performed as described¹⁷ using EdU detection cocktail (Invitrogen). Briefly, seedlings were treated with 1 μM EdU for 30 min and fixed in 4% (w/v) formaldehyde solution in PBS solution with 0.1% Triton X-100 for 30 min. Fixer was washed with PBS (3 \times 10 min) then incubated in EdU detection cocktail for 30 min in the dark, followed by PBS wash (3 \times 10 min) before observation by microscope.

Microscopy and imaging. All images were recorded with a Leica DFC digital camera mounted to a Leica DM5000 microscope using an FITC (fluorescein isothiocyanate)-specific filter (EdU and GFP) or a DIC (differential interference contrast) filter (transparent roots in 89% lactic acid) except for Supplementary Fig. 9a, in which images were collected with an Olympus FV-1000 confocal microscope with a 488 nm Argon laser (GFP).

RT–PCR analyses. Total RNA was isolated from seedlings with TRIzol reagent (Invitrogen) except for the *e2fa* mutant analysis (Fig. 6e), in which total RNA was isolated from the root meristem (0.5 mm root tips). First strand cDNA was synthesized from 1 μg of total RNA with M-MLV (Moloney murine leukemia virus) reverse transcriptase (Promega). All qRT–PCR analyses were performed by CFX96 real time PCR detection system with iQ SYBR green supermix (Bio-Rad). *TUB4* (*At5g44340*) and *EIF4a* (*At3g13920*) were used as control genes.

Microarray analyses. Details for microarray data set sources and data analyses are described in Supplementary Methods. Raw CEL files and RMA log₂ signal intensity files are available at Gene Expression Omnibus (GSE40245).

40. Dello Ioio, R. *et al.* Cytokinins determine *Arabidopsis* root-meristem size by controlling cell differentiation. *Curr. Biol.* **17**, 678–682 (2007).
41. Fu, X. & Harberd, N. P. Auxin promotes *Arabidopsis* root growth by modulating gibberellin response. *Nature* **421**, 740–743 (2003).
42. González-García, M. P. *et al.* Brassinosteroids control meristem size by promoting cell cycle progression in *Arabidopsis* roots. *Development* **138**, 849–859 (2011).

The architecture of *Tetrahymena* telomerase holoenzyme

Jiansen Jiang^{1,2,3*}, Edward J. Miracco^{2*}, Kyungah Hong⁴, Barbara Eckert⁴, Henry Chan², Darian D. Cash², Bosun Min⁴, Z. Hong Zhou^{1,3}, Kathleen Collins⁴ & Juli Feigon^{2,3}

Telomerase adds telomeric repeats to chromosome ends using an internal RNA template and a specialized telomerase reverse transcriptase (TERT), thereby maintaining genome integrity. Little is known about the physical relationships among protein and RNA subunits within a biologically functional holoenzyme. Here we describe the architecture of *Tetrahymena thermophila* telomerase holoenzyme determined by electron microscopy. Six of the seven proteins and the TERT-binding regions of telomerase RNA (TER) have been localized by affinity labelling. Fitting with high-resolution structures reveals the organization of TERT, TER and p65 in the ribonucleoprotein (RNP) catalytic core. p50 has an unanticipated role as a hub between the RNP catalytic core, p75–p19–p45 subcomplex, and the DNA-binding Teb1. A complete *in vitro* holoenzyme reconstitution assigns function to these interactions in processive telomeric repeat synthesis. These studies provide the first view of the extensive network of subunit associations necessary for telomerase holoenzyme assembly and physiological function.

Telomerase is a unique endogenous eukaryotic reverse transcriptase (RT) required for maintenance of linear chromosome ends and is a highly regulated determinant of cellular ageing, stem cell renewal and tumorigenesis^{1,2}. TER contains a region of sequence complementarity to the telomeric repeat that is used as a template for addition of successive repeats to the 3' end of chromosomes, for example addition of the repeat TTAGGG in humans. The specialized telomerase catalytic cycle of telomeric repeat synthesis has been extensively investigated for the human enzyme and biochemically similar enzymes from model organisms such as the telomere-rich ciliate *Tetrahymena thermophila*³. Whereas only TERT and TER are required for telomerase catalytic activity *in vitro*, the physiologically functional holoenzyme is a multi-subunit RNP^{4,5}. Affinity purification, mass spectrometry and subunit tagging assays have identified eight telomerase-specific, reciprocally co-purifying subunits of *Tetrahymena* telomerase holoenzyme, each essential for telomere length maintenance^{6,7} (Fig. 1a). Three of these subunits are required for *in vivo* assembly of a catalytically active RNP: TERT, TER and the La-family protein p65⁷. The p65 La and RRM1 domains increase p65 affinity for TER, but only the carboxy-terminal xRRM2 domain (Fig. 1a) is critical for the TER folding that enhances TERT RNP assembly *in vitro* and *in vivo*^{8–11}. Two TER regions, loop 4 (L4) next to the p65 xRRM2 interaction site and the TERT high-affinity binding element (TBE) 5' to the template (Fig. 1a), bind to the TERT high-affinity RNA binding domain (TRBD)^{12–14}. The TERT TRBD, RT domain and C-terminal extension (CTE) form a ring that encircles an active site cavity, with a fourth TERT amino-terminal (TEN) domain (Fig. 1a) positioned in a TER-dependent but otherwise unknown location^{15–19}.

Little is known about the configuration and roles of the five additional *Tetrahymena* telomerase holoenzyme proteins required for *in vivo* telomere elongation (p75, p50, p45, p19 and Teb1; Fig. 1a); only Teb1 has a known domain structure. Like the paralogous large subunit of the single-stranded DNA (ssDNA) binding factor replication protein A, Teb1 has four OB-fold domains, NABC (Fig. 1a); A and B

bind ssDNA with high affinity and C is necessary for holoenzyme association^{6,20,21}. Teb1 is required for the particularly high DNA product interaction stability of an endogenously assembled *Tetrahymena* telomerase holoenzyme, evident as high repeat addition processivity (RAP) *in vitro*^{6,20,21}. The holoenzyme subunits p75, p19 and p45, here designated 7-1-4, form a subcomplex that remains assembled upon micrococcal nuclease-induced dissociation of the other holoenzyme subunits *in vitro*^{6,21}. Little is known about p50, which seems substoichiometric on silver-stained gels of affinity purified holoenzyme⁶.

Here we report the native electron microscopy structure and a complete *in vitro* reconstitution of *Tetrahymena* telomerase holoenzyme. This first physical and functional network architecture of a telomerase holoenzyme provides unprecedented detail about the structure of the RNP catalytic core and reveals the organization of holoenzyme subunits that confer processivity and bridge telomerase to telomeres.

Overall structure and localization of subunits

Tetrahymena telomerase holoenzymes were prepared for electron microscopy (Supplementary Fig. 1) by affinity purification⁶ from 10 different strains bearing N- or C-terminal 3×Flag (F) and tandem protein A (ZZ) tags (ZZF or FZZ, respectively) on TERT or other holoenzyme subunits (Supplementary Table). Comparison of the class averages from negative staining electron microscopy images of telomerase purified using TERT–FZZ (TERT–F telomerase, which results from cleavage of the ZZ tag) to those from cryo-electron microscopy images (Fig. 1b and Supplementary Fig. 2a, b) shows that the negative stain did not change the structure within experimental resolution. In addition to the predominant ('stable') conformation of the complete holoenzyme particle, constituting 31% of the total particles, the class averages show additional particle subpopulations of varying conformation and subunit composition (described below and Supplementary Fig. 2). The class averages from both negative staining electron microscopy images and cryo-electron microscopy

¹Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, California 90095, USA. ²Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA. ³California Nanosystems Institute, University of California, Los Angeles, California 90095, USA. ⁴Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

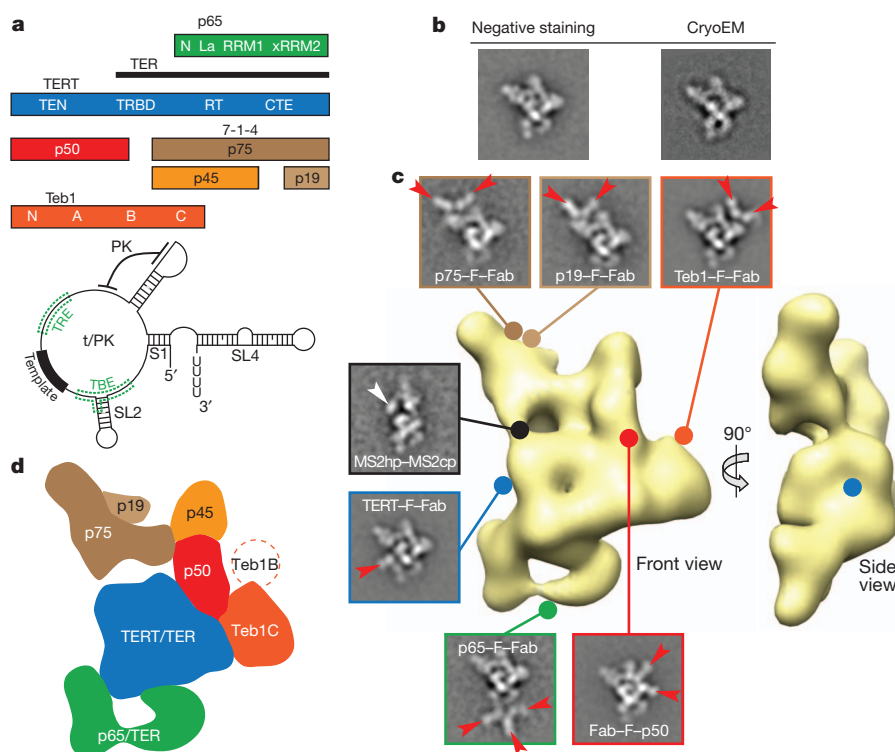


Figure 1 | Electron microscopy reconstruction of *Tetrahymena* telomerase holoenzyme and subunit localization. **a**, Holoenzyme subunits and domains (top) and TER secondary structure (below). **b**, Representative class averages of negative staining electron microscopy and cryo-electron microscopy images of TERT-F telomerase. **c**, 3D reconstruction of Teb1-F telomerase (front and side views) and class averages of affinity-labelled telomerase particles. Lines with circle heads indicate attachment point of Fab (red arrows) and MS2cp (white arrow). Side-lengths of class averages in this and subsequent figures are 350 Å. **d**, Subunit schematic (front view).

images show a strongly preferred orientation (Supplementary Fig. 2). To overcome the problems of preferred orientation and structural variability, the electron microscopy image acquisition and three-dimensional (3D) reconstructions were carried out using an automated random conical tilt (RCT) method²² (Supplementary Fig. 3). The 3D structures of TERT-F and Teb1-F telomerase holoenzymes are indistinguishable (Supplementary Fig. 4a), and Teb1-FZZ gave higher yields of holoenzyme. A 3D structure (Fig. 1c) of the ~500 kDa Teb1-F telomerase holoenzyme from 3D reconstruction of 2,220 particles with a resolution of ~25 Å (Supplementary Fig. 4b) demonstrates well-ordered holoenzyme density occupying about $200 \times 150 \times 80$ Å.

To locate each protein subunit within the overall holoenzyme structure, telomerase was purified from strains harbouring FZZ-tagged TERT, p75, p65, p45, p19 or Teb1 in place of the corresponding untagged protein⁶. During purification, the ZZ portion of the tag was removed by proteolytic cleavage, leaving only the short F tag that was labelled using the antigen-binding fragment (Fab) derived from monoclonal anti-Flag antibody (see Methods). Each tagged protein can bind up to three Fabs, as illustrated in some of the class averages and 3D maps (Fig. 1c and Supplementary Fig. 5), which pinpoint to a single spot and localize the binding site of Fab unambiguously. Affinity purification of the biologically functional C-terminally tagged p50 (p50-FZZ) does not enrich sufficient holoenzyme for silver staining detection in SDS-PAGE, probably due to proteolysis between the C-terminal tag location and the region of p50 required for holoenzyme assembly⁶. Instead we used N-terminally ZZF-tagged p50 (ZZF-p50) expressed in partial replacement of the endogenous locus, which did purify holoenzyme to homogeneity (Supplementary Fig. 6). All holoenzyme protein subunits were localized by Fab labelling except p45, which was identified by process of elimination (see Methods). In addition, we affinity purified telomerase from a strain with biologically functional tagged TER harbouring a small hairpin tag (MS2hp), which is recognized by the MS2 coat protein (MS2cp), appended to TER stem 2 (S2)²³ (Fig. 1c and Supplementary Fig. 5g). The MS2cp dimer bound to this tag appears as extra density in the class averages and thus provides the approximate location of the TBE 5' to the template (Fig. 1a).

The positions of the TERT and Teb1 C termini and the p50 N terminus could be precisely mapped in the 3D reconstructions of

the respective Fab-labelled telomerase particles (Supplementary Fig. 5a–c), whereas a less precise location of the C termini of p65, p75 and p19 (which due to low purification yield did not give sufficient particles for RCT data collection) was obtained from the class averages (Supplementary Table, Fig. 1c and Supplementary Fig. 5d–f). Approximate subunit boundaries (schematized in Fig. 1d) could be modelled after fitting the holoenzyme density with known TERT^{15–18}, TER^{8,24–26}, p65⁸ and Teb1²⁰ high-resolution domain structures combined with comparisons of class averages and 3D reconstructions across different complexes as described below. These data reveal that, in the 'front' view in Fig. 1c (schematized in Fig. 1d) with p65 at the bottom, TERT occupies the centre of the particle and 7-1-4 subunits form the top. Teb1 projects from the middle layer to the right. The p50 subunit is also part of the middle layer, networked between TERT, 7-1-4 and Teb1 (Fig. 1c, d). Numerous inter-subunit appositions generate an intricate and highly contoured surface.

The structure of the RNP catalytic core

Based on the determined subunit locations we fit the crystal and NMR structures of TER and protein domains in the electron microscopy density map of Teb1-F telomerase to generate a model of the entire RNP catalytic core (Fig. 2a and Supplementary Video). The RT and CTE domains of *Tetrahymena* TERT were homology modelled²⁷ from the *Tribolium castaneum* TERT crystal structure¹⁸ and combined with the crystal structures of *Tetrahymena* partial TRBD¹⁵ and TEN¹⁶ domains as described in Methods. The modelled TERT TRBD-RT-CTE fits in only one orientation (Fig. 2a, b and Supplementary Fig. 7a, b). The position of the TERT CTE is consistent with the location of the C terminus identified in the 3D reconstruction of TERT-F-Fab telomerase (Fig. 1c and Supplementary Fig. 5a)¹⁷. The TEN domain was placed into density remaining after determination of subunit boundaries of the adjacent Teb1C and p50, and oriented based on the homology model of human TERT²⁸. The modelled *Tetrahymena* TERT TRBD interacts with the CTE, which is consistent with the *Tribolium* TERT crystal structures^{17,18} and isolated human TERT domain interactions¹⁹. There is electron microscopy density linking the TEN and CTE regions and the TEN and TRBD regions, which could correspond to the ~70 amino acids of potentially unstructured

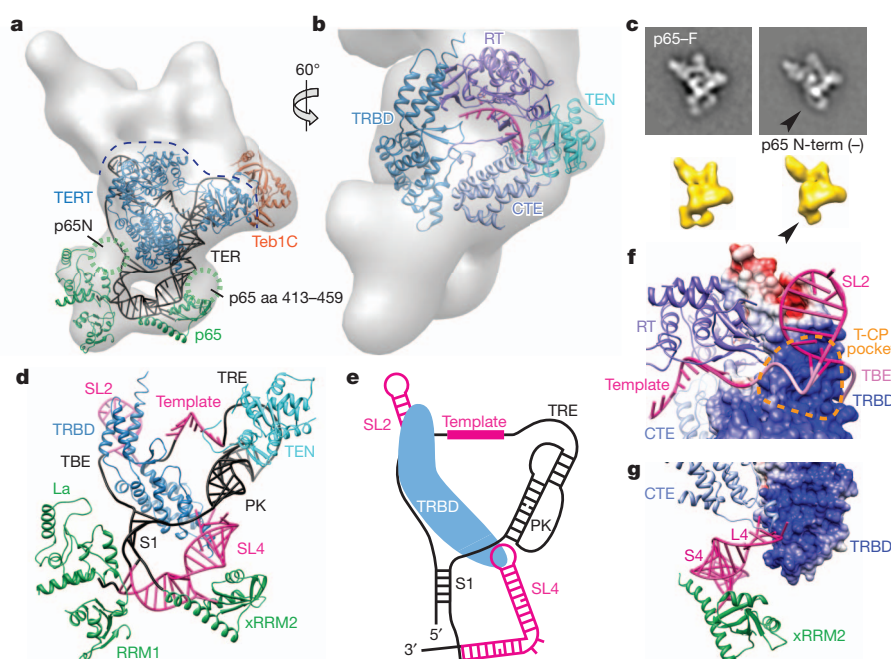


Figure 2 | Structure of the RNP catalytic core. **a**, 3D reconstruction of *Teb1*-F telomerase with TERT, p65 and TER (black), plus *Teb1C* modelled into the electron microscopy density. The dashed line indicates the top boundary of TERT/TER. **b**, Zoomed and rotated view of **a** showing TERT domains TEN, CTE, TRBD and RT, with TER template and essential Mg^{2+} at the active site in magenta. **c**, Class averages of p65-F telomerase (top) and 3D reconstructions (bottom) of TERT-F telomerase with p65 (left) and p65 missing N-terminal density (right, black arrows). **d**, TER model structure (well determined, magenta; remaining, black) and interactions with TERT TRBD and TEN and p65 La, RRM1, and xRRM2 domains. **e**, Secondary structure schematic of TER with TRBD. **f**, Modelled interaction between TBE (pink) and TRBD T-CP pocket. **g**, Modelled interactions of distal SL4 with CTE and bottom of TRBD. In **f**, **g**, TRBD is shown as GRASP surface.

TEN-TRBD linker^{29,30} and adjacent TRBD sequence missing from high-resolution structures^{15,16}, which crosslinks to ssDNA and improves template boundary definition^{19,31}.

The C terminus of p65 is at the bottom of the particle, below TERT (Fig. 1d and Supplementary Fig. 5d). In the class averages and 3D reconstructions we observed particle subpopulations lacking part of the density at the bottom left of the particle (Fig. 2c and Supplementary Figs 2, 3c and 5d), which must arise from loss of p65 N-terminal domain(s) due to partial proteolysis of this subunit common during holoenzyme purification⁶ and/or consequent loss of positional constraints on the remaining La and RRM1 domains. The structure of the p65 C-terminal domain xRRM2-stem 4 (S4) complex⁸ could therefore be localized and fit to the density present in all the 3D reconstructions (Fig. 2a, d and Supplementary Fig. 7c, d). The remaining parts of S1-SL4 and p65 La and RRM1 could then be modelled (see Methods). The p65 N-terminal domain (near S1) and the long $\beta 2$ - $\beta 3$ loop within xRRM2 that was deleted in the crystal structure could occupy the remaining unmodelled density (Fig. 2a)^{8,26}. Together these fittings provide the overall topology of p65-TER interaction (Fig. 2a, d).

Tetrahymena TER contains two major domains, the template/pseudoknot (t/PK) and SL4, which are connected by S1 (Fig. 1a). Starting with the defined locations of the template in the active site of TERT¹⁸, S2 by MS2cp labelling, and SL4 in complex with p65 xRRM2⁸ (Fig. 2d, magenta), and considering topological restrictions based on the length of the single-stranded regions of TER, a model structure of the PK, and S1, we traced a potential trajectory of the entire TER (Fig. 2a, d, e) fit into the remaining electron microscopy density. The template recognition element (TRE) seems to be close to the TEN domain, as implicated biochemically³², and the bottom of the PK is close to L4 (Fig. 2a, d). The locations of TER elements in the model are consistent with the large body of biochemical data on TER structure and function^{4,5,33,34}. Of particular significance are the well-determined locations of the two TRBD binding elements, the TBE and L4^{9,10,13,32}, which we find bind to two distinct regions of the TRBD, near the top left and bottom of the TERT ring, respectively (Fig. 2b, d, e).

Remarkably, only the apical loop of SL4 connects to the density assigned to TERT, resulting in a U-shape at the bottom of the holoenzyme formed by TER S1/SL4 bound by p65. (Fig. 2a, d, e). The interaction of distal SL4 with TERT explains how it can stimulate telomerase activity when added to the t/PK in *trans*^{35,36} and how binding of p65 xRRM2 to S4 mediates hierarchical assembly of p65-TER with TERT⁸⁻¹⁰. The electron microscopy structure and

model reveal that the TRBD bridges the TBE and L4, which are ~ 40 Å apart. In the model, distal SL4 approaches the 3' end of the PK, consistent with reported high fluorescence resonance energy transfer (FRET) between L4 and an internal PK position³⁷. Although detailed interactions await a high-resolution structure, in the model the TBE contacts the TRBD in the T-pocket and CP-motif region (T-CP pocket) (Fig. 2f), which is consistent with biochemical data^{12,14,15}, and distal SL4 is flanked by the CTE and TRBD (Fig. 2g), potentially contributing to folding of TERT around TER.

p50 anchors the accessory proteins

The catalytic activity associated with telomerase purified using ZZFP-p50 has low RAP, in contrast with the high-RAP activity enriched by the much lower yield of holoenzyme purified by p50-FZZ (Fig. 3a). The Fab labelling showed that the N terminus of p50 and the C terminus of *Teb1* are relatively close together (Fig. 1c and Supplementary Fig. 5b, c). Comparison of the class averages of F-p50 and *Teb1*-F telomerase shows that all F-p50 telomerase particles lack density for *Teb1* and reveals the boundary of *Teb1* by difference map (Fig. 3b). The boundary of p50 was further defined by comparing class averages from a small subclass of particles containing only the RNP catalytic core, which was observed only in the MS2cp-bound MS2hp telomerase, with class averages containing p50 (Fig. 3c). Superpositions of 3D reconstructions of Fab-F-p50 and *Teb1*-F telomerase on TERT-F telomerase lacking *Teb1* (see below) show that Fab bound to the N terminus of F-p50 occupies the same space as *Teb1* (Fig. 3d). Together, these results identify the locations of *Teb1* and p50, show that p50 interacts directly with the RNP catalytic core, and reveal that *Teb1* is in close proximity to the N terminus of p50 and to TERT. The holoenzyme structure therefore explains the low-RAP activity of purified N-terminally tagged p50 (Fig. 3a), because enzyme purification by the N-terminal ZZFP-p50 tag disrupts *Teb1* binding. Beyond density assigned to the RNP catalytic core, *Teb1*, and p50, the remaining density in the top part of the 3D map is occupied by the 7-1-4 subcomplex. Taken together, 3D reconstructions reveal a central location for p50 in holoenzyme architecture, as an interaction hub between TERT, 7-1-4 and *Teb1* (schematized in Fig. 3f).

Assigning function to subunit architecture

In parallel with structural analysis, we developed a method for holoenzyme reconstitution from entirely recombinant subunits. RNP catalytic

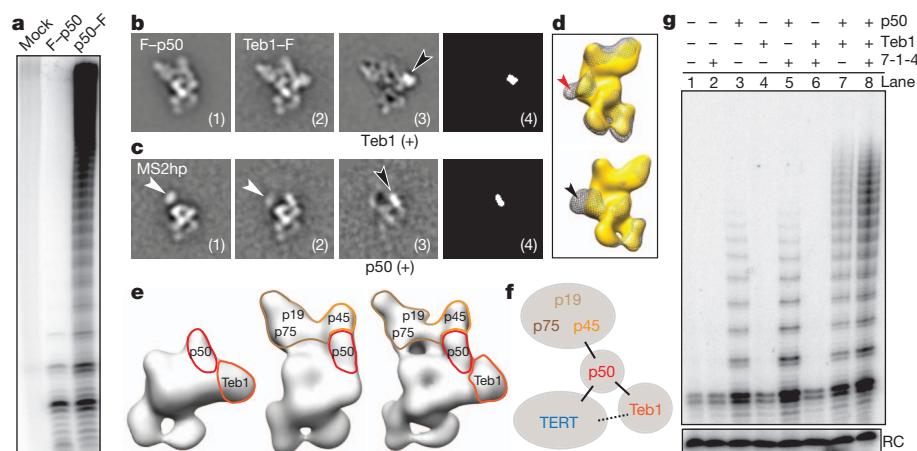


Figure 3 | p50 anchors TERT, 7-1-4 and Teb1. **a**, Primer extension assay of Flag antibody purifications from cell extracts lacking a tagged subunit (mock) or with F-p50 or p50-F. **b**, Class averages of F-p50 telomerase (1), Teb1-F telomerase (2), difference map by subtracting (1) from (2) (seen in 3), and map of statistically significant ($>4\sigma$) regions in the difference map (4). Black arrow points to Teb1 density. **c**, MS2hp telomerase class averages containing MS2cp without p50 (1), with p50 (2), difference map (3), and statistically significant regions (4) as in **b**. White and black arrows point to MS2cp and p50 densities, respectively. 7-1-4 is not seen in these class averages. **d**, Back view of 3D

reconstructions of TERT-F telomerase lacking Teb1 (gold) overlaid with (top) Fab-F-p50 (grey mesh) and (bottom) Teb1-F telomerase (grey mesh) showing that Fab (red arrow) occupies the same site as Teb1 (black arrow). **e**, 3D reconstructions of TERT-F telomerase lacking 7-1-4 (left), lacking Teb1 (middle), and holoenzyme (right). **f**, Schematic of subunit interactions. **g**, Primer extension assay of the RNP catalytic core (TERT-TER-p50) reconstituted with additional combinations of 7-1-4, p50, and/or Teb1BC. Reactions were for 5 min. RC is a radiolabelled oligonucleotide added to telomerase products as a precipitation recovery control.

core with TERT-F was preassembled in rabbit reticulocyte lysate (RRL) and then combined with RRL-expressed p50 and/or 7-1-4 and/or bacterially expressed Teb1BC (see Methods)²¹ (Fig. 3g). Direct primer-extension assays showed that addition of 7-1-4 or Teb1 alone did not alter the low-RAP product synthesis profile of the RNP catalytic core, but addition of p50 alone stimulated enzyme activity and the synthesis of multi-repeat products (Fig. 3g, lane 3). For p50-containing enzymes, 7-1-4 alone increased the amount of product (Fig. 3g, lane 5), Teb1 alone increased product length (lane 7), and their combination into complete holoenzyme was synergistic for increased activity level and long product synthesis (lane 8). These *in vitro* biochemical activities of p50, 7-1-4 and Teb1 were evident within a 5-min reaction time that allowed all product DNAs to be resolved by gel electrophoresis (Fig. 3g) or with the additional time and concentration of dGTP that supported very long product synthesis (Supplementary Fig. 8). The p50-dependent activity of 7-1-4 and Teb1 explains the presence of p50 in virtually all of the electron microscopy structures.

Previous reconstitution assays combined RRL-assembled recombinant RNP catalytic core, bacterially expressed Teb1 and endogenously assembled *Tetrahymena* proteins that remained bound to p45-F after micrococcal nuclease treatment of purified holoenzyme^{20,21,38}. We found that the 7-1-4 complex isolated by this method does not completely dissociate p50, which as a full-length protein co-migrates with p45-F (Supplementary Fig. 6). The presence of residual p50 accounts for the reconstitution of holoenzyme-like catalytic activity by complementation of the RNP catalytic core and Teb1 in previous assays^{20,21,38}. Reconstitutions using entirely recombinant holoenzyme subunits demonstrate a p50-dependent influence of 7-1-4 or Teb1 on RNP catalytic activity (Fig. 3g), consistent with their locations in the electron microscopy structures (Fig. 3e, f).

Teb1 domains are positionally flexible

All telomerase preparations except for those with tagged Teb1 have ~5% of particles missing density for Teb1 in the class averages and 3D reconstructions (Fig. 3e and Supplementary Fig. 2). This is not surprising, because Teb1 is sensitive to proteolysis and dissociation during holoenzyme affinity purification⁶. To identify specific Teb1 domains in the structures, we constructed strains in which TERT was C-terminally tagged at its endogenous locus with only tandem protein A domains (TERT-ZZ) and an N-terminally F-tagged Teb1C

or Teb1BC was expressed from an integrated transgene. Holoenzyme particles isolated by tandem affinity purification of TERT-ZZ/F-Teb1C or TERT-ZZ/F-Teb1BC had the expected high RAP (Fig. 4a) and virtually identical class averages and 3D reconstructions to each other and to those from purified Teb1-F (Fig. 4b-d and Supplementary Fig. 4a). Because Teb1C mediates Teb1 association with other holoenzyme proteins²¹ and the N-terminal tag on Teb1BC absolutely requires the associated holoenzyme to contain Teb1B, we conclude that the Teb1B domain is positionally flexible and therefore did not appear in the class averages and 3D reconstructions. Similarly, for Teb1-FZZ purified particles, which could contain all of the Teb1NABC domains, only Teb1C is visible in the majority of class averages. Teb1C was modelled into the electron microscopy density map (Fig. 2a) based on the position of the C terminus identified by Fab labelling and the best fit with the shape of the electron microscopy density. The orientation around the y axis is not definitive. We observed a small set of class averages (containing $<5\%$ of the particles) of F-Teb1BC and Teb1-F telomerases that show a weak extra density above Teb1C (Fig. 4c, d), which we assign to Teb1B. The conformational flexibility of Teb1 NAB domains relative to Teb1C is consistent with structural studies of the paralogous large subunit of replication protein A, which has three DNA-binding OB-fold domains that become ordered relative to each other only upon binding to ssDNA³⁹.

7-1-4 subcomplex has multiple orientations

Neither high-resolution structures nor exact biological functions of the 7-1-4 subunits are known. Both the electron microscopy structures and previous biochemical data^{6,21} indicate they constitute a stably assembled subcomplex. Because no class averages lacking an individual 7-1-4 subunit were observed, inter-subunit boundaries could only be inferred from the class averages and Fab labelling of the C termini of p19 and p75, which showed that these two subunits are close together. In all class averages and 3D reconstructions, the major class of particles had a 7-1-4 conformation with p75 positioned across the top of the RNP catalytic core (Fig. 1c and Supplementary Fig. 2) in the stable conformation (Fig. 5). All telomerase samples show other positions of 7-1-4 in which it hinges away from the RNP catalytic core while maintaining a physical connection to p50 (Fig. 5 and Supplementary Fig. 3c). In these other conformations, the region of the class averages where 7-1-4 is located appears 'fuzzier'

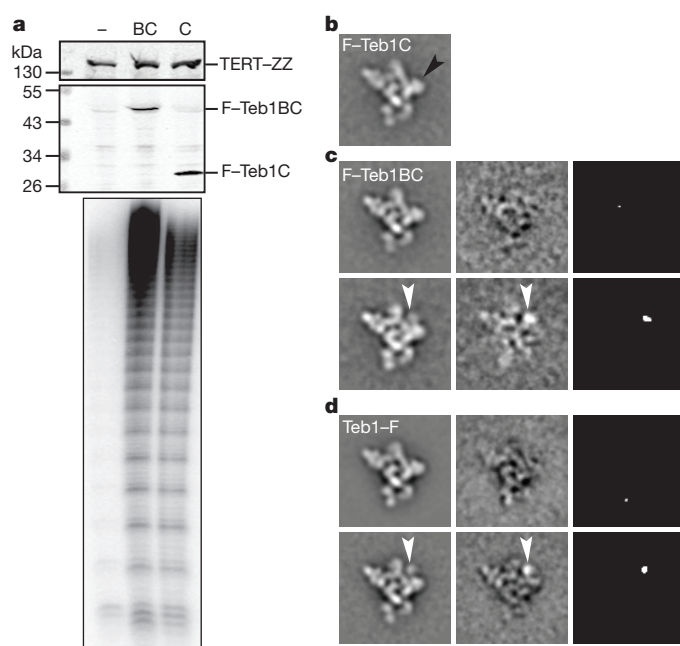


Figure 4 | Contribution of Teb1 domains to holoenzyme structure and activity. **a**, Cell extract western blots and two-step purified enzyme primer extension assays of F-Teb1BC (BC, lane 2) and F-Teb1C (C, lane 3) telomerases. Cell extract with TERT-ZZ alone (–, lane 1) is a negative control for specificity of F-Teb1BC and F-Teb1C binding to Flag antibody. **b–d**, Comparison of class averages (left column) of F-Teb1C (**b**), F-Teb1BC (**c**) and Teb1–F (**d**) telomerases. Density assigned to Teb1B (white arrows) was seen in <5% of particles, whereas density for Teb1C (black arrow in **b**) occupies a fixed position in all particles. Class averages without and with Teb1B density are represented by the upper and lower rows of **c** and **d**, respectively. Difference maps (middle column) by subtracting **b** from the respective class averages and maps of statistically significant (>4 σ) regions in the difference maps (right column) in **c** and **d** show Teb1B density (white arrows).

than the rest of the particle, consistent with conformational variability in this part of telomerase holoenzyme. In ~10% of class averages from every purification, including those with tagged p75 or p45, no 7-1-4 density is visible, indicating that some 7-1-4 positions are not well defined relative to the RNP catalytic core (Fig. 3e and Supplementary Fig. 2). Binding of MS2cp to MS2hp holoenzyme seems to favour p75 displacement from the stable position (Fig. 1c and Supplementary Fig. 5g). Analysis of the various conformations indicates that 7-1-4 is rotating as an intact substructure (Fig. 5). Telomerase product DNA can be ultraviolet crosslinked to p45 as well as the TERT TEN domain^{31,40}, which suggests that at least some orientations of 7-1-4 must bring p45 close to DNA.

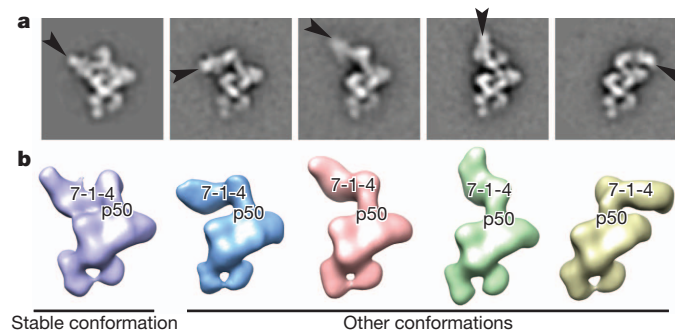


Figure 5 | Positional dynamics of 7-1-4. **a**, Teb1–F telomerase class averages with p75 indicated (black arrows). **b**, 3D reconstructions of Teb1–F telomerase showing different positions of 7-1-4.

Implications for holoenzyme assembly and function

TERT and TER comprise a minimal set of telomerase subunits for repeat synthesis *in vitro*^{5,41}, but despite extensive efforts the 3D arrangement and molecular interactions between TERT and TER have not been previously defined. Although divergent in size and sequence, TERs contain a t/PK domain and almost without exception a separate activating domain (SL4 in *Tetrahymena*), each with a TERT binding site^{33,42}. Here we have unambiguously located two elements of *Tetrahymena* TER that constitute known binding sites for TERT and modelled a path of the entire TER. The functional equivalent of *Tetrahymena* TER distal SL4 in vertebrate TER is P6/P6.1, which ultraviolet crosslinks to the same region of the TRBD where L4 is located in our model⁴³. Our model structure thus provides insight into human telomerase TERT–TER interactions and explains the separable interactions of the TERT binding elements on the t/PK and activating domains of TER.

Using classification and/or 3D reconstruction of 16 differently labelled telomerase samples (Supplementary Table), we were able to identify the subunit arrangement of the holoenzyme. The electron microscopy results together with reconstituted enzyme activity assays suggest a subdivision of holoenzyme functional units: the TERT–TER–p65 catalytic core, p50–Teb1, and 7-1-4. Both structure determination and activity assays of reconstituted holoenzyme subcomplexes establish a central role for p50, a subunit lacking any predicted domain folds. The electron microscopy data reveal that 7-1-4 is a structural unit with dynamic orientation relative to the rest of the holoenzyme. We suggest that the positional dynamics of 7-1-4 coordinate telomerase holoenzyme with additional telomere synthesis or processing activities.

In the model for holoenzyme subunit interactions, the TERT TEN domain, the lower region of p50 and Teb1C have a triangulated arrangement that suggests coordination by pairwise direct contacts. The TERT TEN domain is required for RAP and has been implicated in both ssDNA and TER binding^{5,41,42}. Although Teb1 seems to contact TERT, subcomplex structures never have Teb1 in the absence of p50, and reconstituted enzyme activity assays do not show Teb1 function in the absence of p50. The assembly of p50 with the RNP catalytic core markedly increases processive repeat synthesis, indicating that p50 stabilizes enzyme association with ssDNA in potential functional analogy to yeast Est3 and vertebrate TPP1^{44–48}, favouring a productive TERT TEN domain conformation and/or providing additional ssDNA contact. Subsequent p50-dependent recruitment of Teb1 further enhances activity, which we speculate could occur through interaction of p50 and the Teb1C OB fold. The physical location of p50 determined by electron microscopy and its functional significance determined by reconstitution assays suggest that p50 could be a critical determinant of holoenzyme assembly *in vivo*. This first view of the architecture of a complete telomerase holoenzyme provides unprecedented opportunity to understand the biological mechanisms for coupling of telomerase to its telomere substrates.

METHODS SUMMARY

Tetrahymena strain constructions and steps of tag-based affinity purification were done as described⁶ and as in Methods. To label the tagged subunit for electron microscopy, telomerase particles were first purified using anti-Flag M2 antibody resin then bound to rabbit-IgG resin. The telomerase-bound IgG resin was then incubated with Fab derived from anti-Flag M2 IgG, and elution was effected by protease cleavage. Negatively stained electron microscopy specimens were prepared with fresh telomerase samples, stained with 0.8% uranyl formate, and examined with an FEI Tecnai F20 electron microscope operated at 200 kV. Frozen hydrated specimens were prepared using Quantifoil grids and imaged with an FEI Titan Krios electron microscope operated at 120 kV. The image processing tasks, including image classification and RCT reconstruction, were performed as described in Methods.

Telomerase activity assays were performed at room temperature using purified telomerase complexes on Flag antibody resin with standard *Tetrahymena* holoenzyme reaction conditions using 0.3 μ M ³²P-labelled dGTP. Holoenzyme reconstitution used synthetic genes encoding TERT–F, p75, p65, p50, p45 and p19 for

expression in RRL; TER purified following *in vitro* transcription by T7 RNA polymerase; and N-terminally His₆-tagged Teb1BC purified following bacterial expression²¹.

Full Methods and any associated references are available in the online version of the paper.

Received 22 October 2012; accepted 8 March 2013.

Published online 3 April 2013.

- Artandi, S. E. & DePinho, R. A. Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9–18 (2010).
- Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nature Rev. Genet.* **13**, 693–704 (2012).
- Blackburn, E. H., Greider, C. W. & Szostak, J. W. Telomeres and telomerase: The path from maize, *Tetrahymena* and yeast to human cancer and aging. *Nature Med.* **12**, 1133–1138 (2006).
- Egan, E. D. & Collins, K. Biogenesis of telomerase ribonucleoproteins. *RNA* **18**, 1747–1759 (2012).
- Podlevsky, J. D. & Chen, J. L. It all comes together at the ends: Telomerase structure, function, and biogenesis. *Mutat. Res.* **730**, 3–11 (2012).
- Min, B. & Collins, K. An RPA-related sequence-specific DNA-binding subunit of telomerase holoenzyme is required for elongation processivity and telomere maintenance. *Mol. Cell* **36**, 609–619 (2009).
- Witkin, K. L. & Collins, K. Holoenzyme proteins required for the physiological assembly and activity of telomerase. *Genes Dev.* **18**, 1107–1118 (2004).
- Singh, M. *et al.* Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol. Cell* **47**, 16–26 (2012).
- O'Connor, C. M. & Collins, K. A novel RNA binding domain in *Tetrahymena* telomerase p65 initiates hierarchical assembly of telomerase holoenzyme. *Mol. Cell Biol.* **26**, 2029–2036 (2006).
- Stone, M. D. *et al.* Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**, 458–461 (2007).
- Akiyama, B. M., Loper, J., Najjar, K. & Stone, M. D. The C-terminal domain of *Tetrahymena thermophila* telomerase holoenzyme protein p65 induces multiple structural changes in telomerase RNA. *RNA* **18**, 653–660 (2012).
- Bryan, T. M., Goodrich, K. J. & Cech, T. R. Telomerase RNA bound by protein motifs specific to telomerase reverse transcriptase. *Mol. Cell* **6**, 493–499 (2000).
- Lai, C. K., Mitchell, J. R. & Collins, K. RNA binding domain of telomerase reverse transcriptase. *Mol. Cell Biol.* **21**, 990–1000 (2001).
- Lai, C. K., Miller, M. C. & Collins, K. Template boundary definition in *Tetrahymena* telomerase. *Genes Dev.* **16**, 415–420 (2002).
- Rouda, S. & Skordalakes, E. Structure of the RNA-binding domain of telomerase: Implications for RNA recognition and binding. *Structure* **15**, 1403–1412 (2007).
- Jacobs, S. A., Podell, E. R. & Cech, T. R. Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nature Struct. Mol. Biol.* **13**, 218–225 (2006).
- Gillis, A. J., Schuller, A. P. & Skordalakes, E. Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* **455**, 633–637 (2008).
- Mitchell, M., Gillis, A., Futahashi, M., Fujiwara, H. & Skordalakes, E. Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nature Struct. Mol. Biol.* **17**, 513–518 (2010).
- Robart, A. R. & Collins, K. Human telomerase domain interactions capture DNA for TEN domain-dependent processive elongation. *Mol. Cell* **42**, 308–318 (2011).
- Zeng, Z. *et al.* Structural basis for *Tetrahymena* telomerase processivity factor Teb1 binding to single-stranded telomeric-repeat DNA. *Proc. Natl Acad. Sci. USA* **108**, 20357–20361 (2011).
- Min, B. & Collins, K. Multiple mechanisms for elongation processivity within the reconstituted *Tetrahymena* telomerase holoenzyme. *J. Biol. Chem.* **285**, 16434–16443 (2010).
- Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. *J. Microsc.* **146**, 113–136 (1987).
- Cunningham, D. D. & Collins, K. Biological and biochemical functions of RNA in the *Tetrahymena* telomerase holoenzyme. *Mol. Cell Biol.* **25**, 4442–4454 (2005).
- Chen, Y. *et al.* Structure of stem-loop IV of *Tetrahymena* telomerase RNA. *EMBO J.* **25**, 3156–3166 (2006).
- Richards, R. J., Theimer, C. A., Finger, L. D. & Feigon, J. Structure of the *Tetrahymena thermophila* telomerase RNA helix II template boundary element. *Nucleic Acids Res.* **34**, 816–825 (2006).
- Richards, R. J. *et al.* Structural study of elements of *Tetrahymena* telomerase RNA stem-loop IV domain important for function. *RNA* **12**, 1475–1485 (2006).
- Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
- Steczkiewicz, K. *et al.* Human telomerase model shows the role of the TEN domain in advancing the double helix for the next polymerization step. *Proc. Natl Acad. Sci. USA* **108**, 9443–9448 (2011).
- Friedman, K. L. & Cech, T. R. Essential functions of amino-terminal domains in the yeast telomerase catalytic subunit revealed by selection for viable mutants. *Genes Dev.* **13**, 2863–2874 (1999).
- Armbruster, B. N., Banik, S. S. R., Guo, C., Smith, A. C. & Counter, C. M. N-terminal domains of the human telomerase catalytic subunit required for enzyme activity *in vivo*. *Mol. Cell Biol.* **21**, 7775–7786 (2001).
- Romi, E. *et al.* High-resolution physical and functional mapping of the template adjacent DNA binding site in catalytically active telomerase. *Proc. Natl Acad. Sci. USA* **104**, 8791–8796 (2007).
- O'Connor, C. M., Lai, C. K. & Collins, K. Two purified domains of telomerase reverse transcriptase reconstitute sequence-specific interactions with RNA. *J. Biol. Chem.* **280**, 17533–17539 (2005).
- Theimer, C. A. & Feigon, J. Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.* **16**, 307–318 (2006).
- Cole, D. I. *et al.* New models of *Tetrahymena* telomerase RNA from experimentally derived constraints and modeling. *J. Am. Chem. Soc.* **134**, 20070–20080 (2012).
- Lai, C. K., Miller, M. C. & Collins, K. Roles for RNA in telomerase nucleotide and repeat addition processivity. *Mol. Cell* **11**, 1673–1683 (2003).
- Mason, D. X., Goneska, E. & Greider, C. W. Stem-loop IV of *Tetrahymena* telomerase RNA stimulates processivity in trans. *Mol. Cell Biol.* **23**, 5606–5613 (2003).
- Wu, J. Y., Stone, M. D. & Zhuang, X. A single-molecule assay for telomerase structure-function analysis. *Nucleic Acids Res.* **38**, e16 (2010).
- Eckert, B. & Collins, K. Roles of telomerase reverse transcriptase N-terminal domain in assembly and activity of *Tetrahymena* telomerase holoenzyme. *J. Biol. Chem.* **287**, 12805–12814 (2012).
- Fan, J. & Pavletich, N. P. Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev.* **26**, 2337–2347 (2012).
- Rosenfeld, K. K., Ziv, T., Goldin, S., Glaser, F. & Manor, H. Mapping of DNA binding sites in the *Tetrahymena* telomerase holoenzyme proteins by UV cross-linking and mass spectrometry. *J. Mol. Biol.* **410**, 77–92 (2011).
- Collins, K. Single-stranded DNA repeat synthesis by telomerase. *Curr. Opin. Chem. Biol.* **15**, 643–648 (2011).
- Blackburn, E. H. & Collins, K. Telomerase: An RNP enzyme synthesizes DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003558 (2011).
- Bley, C. J. *et al.* RNA–protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl Acad. Sci. USA* **108**, 20333–20338 (2011).
- Lee, J., Mandell, E. K., Rao, T., Wuttke, D. S. & Lundblad, V. Investigating the role of the Est3 protein in yeast telomere replication. *Nucleic Acids Res.* **38**, 2279–2290 (2010).
- Yen, W. F., Chico, L., Lei, M. & Lue, N. F. Telomerase regulatory subunit Est3 in two *Candida* species physically interacts with the TEN domain of TERT and telomeric DNA. *Proc. Natl Acad. Sci. USA* **108**, 20370–20375 (2011).
- Talley, J. M., DeZwaan, D. C., Maness, L. D., Freeman, B. C. & Friedman, K. L. Stimulation of yeast telomerase activity by the ever shorter telomere 3 (Est3) subunit is dependent on direct interaction with the catalytic protein Est2. *J. Biol. Chem.* **286**, 26431–26439 (2011).
- Zaug, A. J., Podell, E. R., Nandakumar, J. & Cech, T. R. Functional interaction between telomere protein TPP1 and telomerase. *Genes Dev.* **24**, 613–622 (2010).
- Sexton, A. N., Youmans, D. T. & Collins, K. Specificity requirements for human telomere protein interaction with telomerase holoenzyme. *J. Biol. Chem.* **287**, 34455–34464 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grants from NSF MCB1022379 and NIH GM48123 to J.F., NIH GM54198 to K.C., GM071940 and AI069015 to Z.H.Z., Ruth L. Kirschstein NRSA postdoctoral fellowship GM101874 to E.J.M., and Ruth L. Kirschstein NRSA pre-doctoral training grant GM007185 fellowship for H.C. and D.D.C. We acknowledge the use of instruments at the Electron Imaging Center for NanoMachines supported by NIH (1S10RR23057 to ZHZ) and CNSI at UCLA.

Author Contributions J.J. and E.J.M. purified and characterized electron microscopy samples, collected and analysed electron microscopy data, and wrote the paper; K.H., B.E. and B.M. designed and made strains, expression plasmids, initial purifications and reconstituted holoenzyme; H.C. purified telomerase; D.D.C. refined and modelled elements of TER; Z.H.Z., K.C. and J.F. supervised the research, analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F. (feigon@mbi.ucla.edu), K.C. (kcollins@berkeley.edu) or Z.H.Z. (hong.zhou@ucla.edu).

METHODS

Strain construction, growth and analysis. *Tetrahymena* strains TERT-FZZ, Teb1-FZZ, p75-FZZ, p65-FZZ, p50-FZZ, MS2hp, p45-FZZ and p19-FZZ were previously described^{6,23}. New strains ZZ-F-p50, TERT-ZZ-F-Teb1C and TERT-ZZ-F-Teb1BC were selected for targeted replacement of the endogenous locus (ZZ-F-p50 or TERT-ZZ) and/or for *MTT1* promoter-driven transgene integration at *BTU1* (F-Teb1C, F-Teb1BC) using the *neo2* and *bsr2* cassettes^{6,49}. Cells were grown to mid-log phase ($2-4 \times 10^5$ cells per ml) at 30 °C in modified Neff medium (0.25% proteose peptone (EMD Milipore Chemicals), 0.25% BD Bacto yeast extract (Becton Dickinson), 0.2% glucose, 30 μ M FeCl₃). For large-scale purifications, transgene expression was induced by addition of 6 μ M CdCl₂ 1 h before harvesting. Cell extract preparation and affinity purifications for subunit identification and activity assays were performed as described⁶. Genotypes were verified by Southern blots and western blots using rabbit IgG (Sigma) to detect the ZZ tag or anti-Flag M2 mouse monoclonal antibody (Sigma) to detect the Flag tag. Reconstituted enzyme activity assays were performed at room temperature using purified telomerase complexes on Flag antibody resin following standard *Tetrahymena* holoenzyme reaction conditions (50 mM Tris-acetate, pH 8.0, 10 mM spermidine, 5 mM 2-mercaptoethanol, 2 mM MgCl₂, 100 nM d(GT₂G₃)₃, 200 μ M dTTP, and 0.3 μ M ³²P-labelled dGTP) for 5 min or with 3.0 μ M ³²P-labelled dGTP for 15 min⁶. Holoenzyme reconstitution used synthetic genes encoding TERT-FZZ, p75, p65, p50, p45 and p19 for expression in RRL; TERT purified following *in vitro* transcription by T7 RNA polymerase; and N-terminally His₆-tagged Teb1BC purified following bacterial expression²¹.

Cell extract preparation and affinity purification for electron microscopy. Cells were collected by centrifugation (5 min, 2,100g), washed in 20 mM HEPES NaOH, pH 8.0, and then resuspended in lysis buffer (30 ml per l of growth) containing H2EG50 (20 mM HEPES NaOH, pH 8.0, 50 mM NaCl, 1 mM EDTA TS, 1 mM TCEP HCl, and 10% glycerol) supplemented with 0.1% Triton X-100, 0.2% IGEPAL CA-630, 0.1 mM PMSF, 5 μ M proteasome inhibitor MG-132, and 1,000-fold dilution of Sigma Protease Inhibitor Cocktail P8340. To effect cell lysis, resuspensions were rotated end-over-end for 20 min and cleared by ultracentrifugation (1 h, 145,000g).

To minimize proteolysis, all purification steps were carried out at 4 °C and additional washes were used in the early steps of purification as described below. Rabbit-IgG agarose slurry (Sigma) (3 μ l per ml of cell extract) was washed 3 \times in lysis buffer, then added to cell extract for binding with end-over-end rotation overnight. Resin was collected by centrifugation (1 min, 3,200g) and washed 3 \times with no incubation and then 3 \times with end-over-end rotation (10 min) in wash buffer (20 mM HEPES NaOH, pH 8.0, 50 mM NaCl, 1 mM MgCl₂, 1 mM TCEP HCl, 10% glycerol, and 0.1% IGEPAL CA-630). Telomerase was eluted with tobacco etch virus (TEV) protease (30 nM) in wash buffer for 1 h. After elution, supernatant containing holoenzyme was incubated with washed anti-Flag M2 affinity gel (Sigma) for 1 h (4 μ l slurry per g of pelleted cells). After binding, anti-Flag M2 affinity gel was washed as above. Elution was effected by addition of glycerol-free wash buffer containing 3 \times Flag peptide (Sigma, 200 ng μ l⁻¹).

Fab labelling for subunit identification. Anti-Flag Fab was prepared by incubating monoclonal mouse anti-Flag M2 IgG (Sigma) with resin-immobilized papain (Thermo Scientific Pierce) and purified according to the manufacturer's protocol. To prevent masking of the 3 \times Flag epitope present on the target subunit by the 3 \times Flag peptide required for elution, the tandem purification was done in reverse order with anti-Flag M2 affinity gel purification followed by rabbit-IgG agarose purification. Washing and elution were done as described above. To effect Fab labelling, holoenzyme-enriched rabbit-IgG agarose was incubated (1 h, 4 °C) with wash buffer containing anti-Flag Fab (7.5 μ g ml⁻¹). Excess Fab was removed by washing 5 \times with wash buffer. Labelled holoenzyme was eluted with wash buffer containing His₆-TEV protease (2 nM). After elution, His₆-TEV protease was removed from the elution by incubating with washed Ni-NTA agarose (30 min, 4 °C).

Holoenzyme purification by MS2cp. *Tetrahymena* expressing MS2hp TER were collected, lysed and clarified as described above. N-terminal His₆-ZZF-MS2cp bacterially expressed from pET28 was purified using Ni-NTA agarose. Rabbit-IgG agarose was incubated with His₆-ZZF-MS2cp in a 1:1 stoichiometry to anticipated holoenzyme yield based on previous preparations. The enriched rabbit-IgG agarose was then washed 3 \times and added to clarified extract for overnight incubation. Purification then proceeded as described above.

Electron microscopy specimen preparation and data collection. For negative staining electron microscopy, 2 μ l of telomerase sample was applied to a glow-discharged grid coated with carbon film. The sample was left on the carbon film for 10 s, followed by negative staining with 0.8% uranyl formate. Negatively stained specimens were carefully examined by electron microscopy, and only the regions with particles fully embedded in stain were selected for imaging. Electron microscopy micrographs were recorded on a TIETZ F415MP 16-megapixel CCD

camera at 68,027 \times magnification in an FEI Tecnai F20 electron microscope operated at 200 kV. The micrographs were saved by 2 \times binning and had a pixel size of 4.4 Å. To overcome the problems of preferred orientation and structural variability of the samples, 3D reconstructions were carried out using RCT method, which takes advantage of classification to differentiate different conformations and tilting to get lateral views for 3D reconstruction²². RCT images were collected with the grids tilted at two angles successively (65° and 0°) for each specimen area of interest with the assistance of Legimon automation software^{50,51}. To minimize the potential smearing/stretching on 3D maps due to the missing cone problem in RCT data collection, a high tilt angle (65°) was used. For samples where only 2D image analysis was performed, micrographs were taken without tilting of the grids. Total numbers of micrographs and particles used for analysis are summarized in the Supplementary Table.

For cryo-electron microscopy, 2.5 μ l of sample was applied to a glow-discharged Quantifoil R2/1 grid. The grid was then blotted with filter paper to remove excess sample, and flash-frozen in liquid ethane with an FEI Vitrobot Mark IV. The grid was loaded into an FEI Titan Krios electron microscope operated at 120 kV. Micrographs were acquired with a Gatan UltraScan4000 16-megapixel CCD camera at 65,162 \times magnification with defocus values ranging from -1.9 to -4.0 μ m and an exposure dose of 20 e⁻ Å⁻², and without tilting of the grids. The micrographs with a pixel size of 2.3 Å were used for data processing without binning.

2D image reference-free classification. Particles from the untilted micrographs were automatically picked using DoGpicker⁵². All particles in the micrographs, including a small number of aggregated particles, were initially picked to avoid bias in selecting particles. Particles were boxed out in 96 \times 96 pixels (or 144 \times 144 pixels for cryo-electron microscopy images) using *batchboxer* in EMAN⁵³. The defocus value of each micrograph was determined by CTFFIND⁵⁴ and the particles were corrected for contrast transfer function (CTF) by phase-flipping with the corresponding defocus and astigmatism values using *bsoft*⁵⁵. The phase-corrected particles were then iteratively (normally 9 iterations) classified using the *refine2d.py* program in EMAN. The classification for each sample was done independently without using any model or reference. The aggregated particles clustered in the few classes showing poor structures were manually eliminated during the iterations of classification. For the samples at very low concentration (Fab-labelled p75-F, p19-F and p65-F telomerase, and MS2cp-labelled MS2hp telomerase), a looser criterion for DoGpicker was used to pick more particles and consequently some 'fake' particles from the stain background were also selected. The class averages were visually inspected at several different iterations of classification to remove the 'fake' particles that were clustered in a small set of class averages that show no structural features.

For the RCT 3D reconstructions (Supplementary Table), the screened particles from the above steps were re-classified using the Correspondence Analysis method in SPIDER⁵⁶. The class averages from the above EMAN classification step were used as references for multi-reference alignment in SPIDER. Each new class from this analysis was used for an RCT 3D reconstruction.

RCT 3D reconstruction and refinement. Corresponding particles from tilt pairs (65° and 0°) were picked using *ApTiltPicker.py* in Appion⁵² modified to optimize automatic particle picking and verified by visual inspection. The 0° tilt particles were classified as described above. Next, 3D RCT maps were reconstructed from 65° tilt particles. Particles within each class of the 0° tilt data set all have the same conformation and the same orientation except in-plane rotation. Their corresponding 65° tilt particles have different orientations due to in-plane rotation and were combined to reconstruct one 3D map. Each 3D map was iteratively refined with SPIDER followed by FREALIGN⁵⁷ by refinement of the centre of 65° tilt particles in the corresponding class. A representative refinement procedure is shown in Supplementary Fig. 3b. The orientations of 65° tilt particles were calculated from the in-plane rotation of the corresponding 0° tilt particles and the geometric relationship between the 65° and 0° tilt micrographs, and kept fixed during structure refinement. Multiple 3D maps were generated from RCT 3D reconstruction because particles in different conformations were classified into different classes each of which generated an individual 3D map. Representative 3D maps are shown in Supplementary Fig. 3c. Each 3D map was refined using only the 65° tilt particles in its corresponding class. The defocus value of each 65° tilt particle was calculated by CTFFILT program⁵⁴ and then applied to 3D reconstructions with FREALIGN for proper correction of phase and amplitude in CTF.

To estimate the resolution of the representative 3D map of Teb1-F telomerase in the 'stable' conformation, the 2,220 65° tilt particles used to reconstruct that 3D map were split into odd- and even-numbered halves and the halves of data were used to reconstruct two 3D maps with FREALIGN using the determined orientations and image centres, respectively, to calculate Fourier shell correlation (FSC). The resolution is 25 Å at FSC = 0.5 (Supplementary Fig. 4b). For comparison, the FSC was also calculated using the recently introduced 'gold standard' FSC

method⁵⁸, in which the halves of data (both 65° and 0° tilt) were split at the beginning for independent orientation determination, 3D reconstruction, and refinement to generate two independent 3D maps. The ‘gold standard’ FSC suggests resolutions of 29 Å at FSC = 0.5 and 24 Å at FSC = 0.143 (Supplementary Fig. 4b). Since a high tilt angle (65°) was used for RCT 3D reconstructions, the missing cone was minimized and did not have notable impact on the structural interpretation of the 3D maps at the present resolution.

Evaluation of sample integrity by size-exclusion chromatography (SEC). To confirm that the affinity-purified telomerase holoenzymes were intact after purification and that negative staining did not induce significant aggregation, we obtained an SEC profile of a sample of affinity purified TERT-F telomerase and electron microscopy images from the peak fractions (Supplementary Fig. 1c, d). The sample (50 µl) was run at 50 µl min⁻¹ on a Superdex 200 column in the same buffer used for the electron microscopy samples except IGEAL CA-630 was replaced by Tween-80. The SEC profile shows that the telomerase holoenzyme runs as a single peak with a constant $A_{260\text{ nm}}/A_{280\text{ nm}}$ ratio indicative of a largely intact RNA-protein complex. The class averages of particles from the peak fraction are identical to those for the affinity-purified sample (data not shown). There are no later eluting peaks from dissociated subunits (which would have different $A_{260\text{ nm}}/A_{280\text{ nm}}$ ratios).

Localization of Fab-labelled subunits. Each protein was assigned by identifying the characteristic Fab density in the class averages and 3D reconstructions. The 3×Flag tag allows up to three Fabs to bind simultaneously. Fab was observed attached to all tagged telomerase particles except for p45-F. For p45-F telomerase, Fabs were observed bound to free p45-F protein, while the p45-F telomerase holoenzyme particles appeared unchanged. The location of p45 in the holoenzyme was therefore assigned based on the density remaining after assignment of TERT-F, Teb1-F, p75-F, p65-F, F-p50 and p19-F.

Fitting of atomic models into the 3D electron microscopy maps. The homology model of *Tetrahymena* TERT RT-CTE domains was built using amino acids 520–1117 and the *Tribolium* TERT crystal structure (PDB ID: 3KYL) as a template with SWISS-MODEL²⁷. The model of *Tetrahymena* TERT (TRBD, RT, and CTE domains) was built by joining the crystal structure of *Tetrahymena* TRBD (PDB ID: 2R4G) and the homology model of RT-CTE domains using the *Tribolium* TERT crystal structure as reference for structural alignment. The model of *Tetrahymena* TERT was then manually placed in an approximate position in the electron microscopy density map, followed by docking into the electron microscopy density map accurately using the “Fit in Map” function in UCSF Chimera⁵⁹. The location of the TEN domain, absent in *Tribolium* TERT¹⁷, was supported by definition of the subunit boundaries of the adjacent Teb1C and p50. The crystal structure of *Tetrahymena* TERT TEN domain (PDB ID: 2B2A) was manually placed into the electron microscopy density map. Its position was iteratively adjusted to fit the empty density adjacent to the boundaries of Teb1 and p50 and the rest of TERT. Because the orientation of the TEN domain could not be determined based on the electron microscopy density due to the limit of resolution, we oriented it to be consistent with the homology model of human TERT²⁸. For fitting of p65 xRRM2:SL4, the putative electron microscopy density of p65-TERT cut from the full electron microscopy density map was used in order to avoid interference from unrelated structures. The model of p65 xRRM2:SL4⁸ was manually placed in the electron microscopy density map based on the location of the C terminus determined by Fab labelling and then docked using the “Fit in Map” function in UCSF Chimera. Its position was further manually adjusted to avoid clashing with the model of TERT. To confirm the fittings of TERT and p65 xRRM2:SL4, these two structures were also fit into the electron microscopy density map using the *colores* program in Situs⁶⁰. The fitting with *colores* was automated by 6D-space search and contour-based matching, and did not require manual placing of high-resolution structures to approximate positions in electron microscopy density maps. The crystal structure of Teb1C (PDB ID: 3U50) was manually placed into the electron microscopy density map and its orientation was adjusted so that the position of its C terminus was consistent with the result from the Fab labelling experiment. The homology model of p65 La-RRM1:AUUUU-3’^{8,61} was manually placed into the electron microscopy density map, and its orientation was adjusted to match the shape of the electron microscopy density map. Cross-correlation for the fittings were: TERT TRBD-RT-CTE, 0.91; TERT TEN, 0.84; p65 xRRM2:SL4, 0.80; p65 La-RRM1:AUUUU-3’, 0.86; and Teb1C, 0.82. Although the position and orientation of p65 xRRM2:SL4 are well determined, the cross-correlation coefficient reflects the absence of the β2–β3 loop and amino acids linking xRRM2 to RRM1 in the crystal structure⁸.

Structures of SL2, distal SL4 and the PK. Previously reported NMR structures of SL2 and distal SL4^{25,26} were re-refined using a new nuclear Overhauser enhancement (NOE) restraint list derived from re-analysis of all previously obtained 2D nuclear Overhauser enhancement spectroscopy (NOESY) and 3D NOESY-heteronuclear multiple quantum correlation (HMQC) spectra and a new set of residual dipolar

couplings (RDC) restraints (63 for SL2 and 48 for distal SL4) collected using pfl bacteriophage for the alignment media. The new PDB accession numbers for SL2 and distal SL4 are 2M22 and 2M21, respectively, and BMRB accession numbers are 18892 and 18891, respectively. A grid search produced optimal values for the magnitude and asymmetry of the RDC alignment tensor of $D_a = -22.3\text{ Hz}$, $R = 0.38$ for SL2, and $D_a = -17.5\text{ Hz}$, $R = 0.40$ for SL4. The structures were calculated from an extended, unfolded RNA conformation using XPLOR-NIH v.2.9.8^{62,63} following standard XPLOR protocols. A model structure of the PK was generated based on the secondary structure and two U-A-U triples predicted by sequence analysis⁶⁴. XPLOR-NIH was used to calculate the structure from an extended RNA conformation as previously described^{62,63}. A mock NMR restraint list was derived based on the solution structure of the human telomerase PK⁶⁵. For the two stems, A-form RNA restraints were used for the dihedral angles and NOEs. NOEs for residues in the stems include H6/H8 (Py/Pu) to its own sugar protons and to the sugar protons (H1’, H2’, H3’, H4’) of the $n - 1$ residue, H6/H8 to H6/H8 sequential, and imino sequential NOEs. For the PK loop 1, base-triple hydrogen bond restraints, planarity restraints with PK stem 2 residues, and sequential stacking NOEs were used to form the structured triple helix. PK loop 2 was given H6/H8 sequential NOEs and H6/H8-H1’ NOEs within the minor groove of PK stem 1 to place it in the expected conformation.

Modelling of TER. First, the position of the template in TER was determined by matching the crystal structure of *Tribolium* TERT-nucleic acid complex (PDB ID: 3KYL) with the docked model of *Tetrahymena* TERT. Second, the position of SL4 was determined by the docking of the atomic model of p65 xRRM2:SL4 complex into the electron microscopy density map. Third, the atomic model of SL2 was manually placed into the electron microscopy density map and its position was iteratively adjusted to match the result from MS2cp labelling experiment. S1 (modelled as A-form RNA) and the PK were placed into the electron microscopy density map based on physical constraints with the rest of TER. S1 is connected to S4 by 4 nucleotides and to S2 by 10 nucleotides. The model of the 31 nucleotide PK, which is connected to S1 by 3 nucleotides and to the template by the TRE, was fit into the only region of unoccupied density, between the TRBD and TEN domains, remaining after all the other major elements of TERT and TER were localized. Although the modelled positions of the TRE 3’ to the template (Fig. 1a) and the top of the pseudoknot could potentially be swapped with the TEN domain, they would still be on the same side (right side or behind in Fig. 2a or 2b, respectively) of the TERT TRBD-RT-CTE ring that has the active site and bound template. In either case, the TRE appears to be close to the TEN domain. The single-stranded regions of TER (except the template) were initially modelled as ideal A-form single-stranded RNA and their structures were modified using the “Minimize structure” function in UCSF Chimera to connect with their respective folded RNA fragments. The electron microscopy density maps and the modelled protein and RNA structures were visualized with UCSF Chimera.

49. Couvillion, M. T. & Collins, K. Biochemical approaches including the design and use of strains expressing epitope-tagged proteins. *Methods Cell Biol.* **109**, 347–355 (2012).
50. Suloway, C. *et al.* Automated molecular microscopy: The new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
51. Suloway, C. *et al.* Fully automated, sequential tilt-series acquisition with Leginon. *J. Struct. Biol.* **167**, 11–18 (2009).
52. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
53. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
54. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
55. Heymann, J. B. Bsoft: Image and molecular processing in electron microscopy. *J. Struct. Biol.* **133**, 156–169 (2001).
56. Frank, J. *et al.* SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
57. Grigorieff, N. FREALIGN: High-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117–125 (2007).
58. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
59. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
60. Chacón, P. & Grigorieff, N. Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.* **317**, 375–384 (2002).
61. Jacks, A. *et al.* Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure* **11**, 833–843 (2003).
62. Schwieters, C. D., Kuszewski, J. J. & Marius Clore, G. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62 (2006).
63. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Marius Clore, G. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).

64. Ulyanov, N. B., Shefer, K., James, T. L. & Tzfati, Y. Pseudoknot structures with conserved base triples in telomerase RNAs of ciliates. *Nucleic Acids Res.* **35**, 6150–6160 (2007).
65. Theimer, C. A., Blois, C. A. & Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* **17**, 671–682 (2005).

The domination of Saturn's low-latitude ionosphere by ring 'rain'

J. O'Donoghue¹, T. S. Stallard¹, H. Melin¹, G. H. Jones^{2,3}, S. W. H. Cowley¹, S. Miller^{3,4}, K. H. Baines⁵ & J. S. D. Blake¹

Saturn's ionosphere is produced when the otherwise neutral atmosphere is exposed to a flow of energetic charged particles or solar radiation¹. At low latitudes the solar radiation should result in a weak planet-wide glow in the infrared, corresponding to the planet's uniform illumination by the Sun². The observed electron density of the low-latitude ionosphere, however, is lower and its temperature higher than predicted by models^{3–5}. A planet-to-ring magnetic connection has been previously suggested, in which an influx of water from the rings could explain the lower-than-expected electron densities in Saturn's atmosphere^{6–8}. Here we report the detection of a pattern of features, extending across a broad latitude band from 25 to 60 degrees, that is superposed on the lower-latitude background glow, with peaks in emission that map along the planet's magnetic field lines to gaps in Saturn's rings. This pattern implies the transfer of charged species derived from water from the ring-plane to the ionosphere, an influx on a global scale, flooding between 30 to 43 per cent of the surface of Saturn's upper atmosphere. This ring 'rain' is important in modulating ionospheric emissions and suppressing electron densities.

On 17 April 2011 over two hours of Saturn near-infrared spectral data were obtained by the 10-m W. M. Keck II telescope using the NIRSPEC (Near InfraRed Spectrograph) spectrometer⁹. The slit on the spectrometer was positioned along Saturn's noon meridian as shown in Fig. 1. The intensity of two bright H_3^+ rotational-vibrational emission lines is almost completely visible from pole to pole, so that low-latitude emissions can be studied. Far from being featureless (as we might expect by analogy to Jupiter¹; see Supplementary Information), the mid- to low-latitude H_3^+ emissions show a number of peaks and

troughs before increasing strongly towards the two polar auroral regions; a number of these peaks are observed in both spectral lines. The $\text{Q}(1, 0^-)$ line shows more substantial peaks and troughs at mid- to low latitudes than does the $\text{R}(2, 2^-)$ line, owing to less contamination by reflected sunlight in neighbouring spectral pixels where methane is not absorbing this light effectively. The apparently symmetric peaks and troughs do not occur at the same latitudes either side of the equator, however: they occur at higher latitudes in the north than in the south. The lack of latitudinal symmetry, along with the absence of a variability similar to Jupiter's¹, suggests that the phenomenon is unrelated to weather patterns or other processes produced in the neutral atmosphere.

Instead, the peaks in emission are found to be mapped via planetary magnetic field lines to gaps in Saturn's rings such as the Cassini division (in which we will include here the Herschel, Laplace, Encke and Keeler gaps) and the Colombo gap. In addition, we define the 'instability region' as the region between the inner edge of the B ring and the 'instability radius'. The inner edge of the B ring and the 'instability radius' are regions within which the outward centrifugal forces on particles balance with the inward gravitational forces within the rings, such that particles are unstable and can easily stream along magnetic field lines and enter Saturn's atmosphere^{10,11}. The model used for this mapping uses the most recent internal field coefficients determined from Cassini data¹², together with small field perturbations produced by magnetospheric currents (see also the Supplementary Information).

The field lines within the equatorial region, where the H_3^+ emission features are found, map to Saturn's main ring system between $1.2R_S$ and $2.3R_S$ from the centre of the planet (where R_S is the 1-bar Saturn

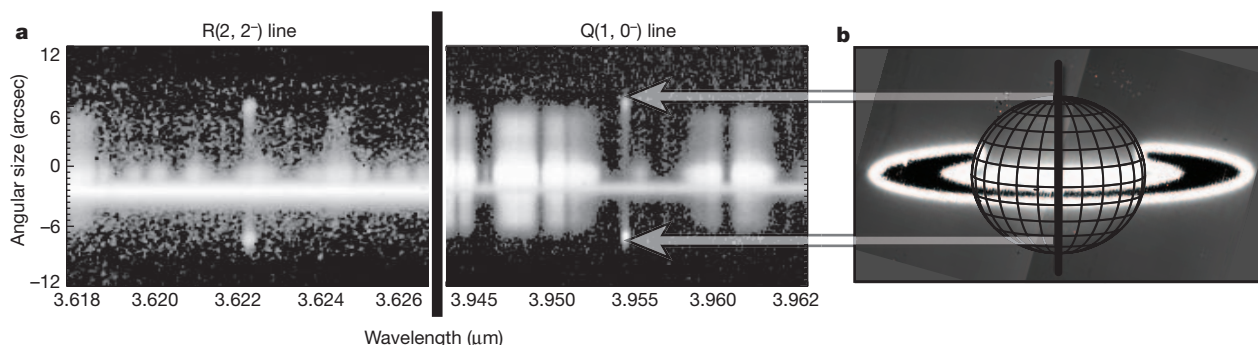


Figure 1 | The process of data acquisition. The spectral images shown in **a** are separated by a thick black line to indicate the different wavelength ranges. The horizontal and vertical axes in these ranges show wavelength and spatial position along the slit, respectively, while intensity ranges from low (black) to high (white). Two spectral lines, $\text{Q}(1, 0^-)$ at $3.953 \mu\text{m}$ and $\text{R}(2, 2^-)$ at $3.622 \mu\text{m}$, are centred in each wavelength range, and are from the Q and R branches of the H_3^+ emission spectrum. These spectra are obtained through the slit of the spectrometer seen in **b**, which was orientated in the north-south position on Saturn, aligned along the rotational axis. Saturn's spin axis was tilted by 8.2°

towards Earth during conditions of Saturn's early northern spring. The planet rotates beneath the slit, allowing the acquisition of spectral images at a fixed local time, but varying in Saturn longitude. In the approximately two hours of recorded data, 21% of the longitude of the planet was observed: 101° – 177° longitude. The bright infrared emission measured at the -3 arcsecond position in **a** across the entire wavelength range is the uniform reflection of sunlight by the rings, and the remaining bright white areas are due to methane reflection. This background methane reflection attenuates the $\text{R}(2, 2^-)$ line emission more than that of the $\text{Q}(1, 0^-)$ line, leading to a lower signal-to-noise ratio.

¹Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. ²Mullard Space Science Laboratory, University College London (UCL), Holmbury St Mary, Dorking, Surrey RH5 6NT, UK. ³The Centre for Planetary Sciences at University College London/Birkbeck, Gower Street, London WC1E 6BT, UK. ⁴Atmospheric Physics Laboratory, Department of Physics and Astronomy, UCL, Gower Street, London WC1E 6BT, UK. ⁵Jet Propulsion Laboratory, California Institute of Technology, MS 183-601, 4800 Oak Grove Drive, Pasadena, California 91109, USA.

equatorial radius of 60,268 km), as shown by the mapped equatorial distances in Fig. 2. A majority of the emission peaks correspond to prominent gaps in the rings, whereas the troughs map to the dense sections of the rings. This relationship is seen in greater spatial detail in Fig. 3, where ring transparency (as measured by Voyager 2) is compared with the total average H_3^+ intensity from the co-addition of both hemispheres and spectral lines.

Cassini spacecraft observations over Saturn's rings during its orbit insertion manoeuvre in 2004, unique in the mission to date, indicate the presence of a water-product atmosphere surrounding the rings: this atmosphere is composed of icy grains and is partly ionized by solar ultraviolet radiation, analogously to the planetary ionosphere^{13,14}. Given the correlation between Saturn's ionosphere and magnetically mapped locations in the rings, we propose that it is this charged material that causes the pattern of features seen on the planet. Water-related ions and electrons must be driven to stream along the field lines into the planetary ionosphere, and we now discuss how this could produce the modulation seen in the ionosphere.

A magnetic link between Saturn's rings and its atmosphere has previously been invoked to explain the lower-than-expected electron densities and their latitudinal variations in the planetary ionosphere, through the influx of water^{6–8}. These earlier observations show discrete dark bands in the lower atmosphere (beneath the ionosphere) taken by the Voyager 2 spacecraft⁷, which the authors interpreted to be magnetically mapped sequentially to the inner edge of the B ring, the instability radius and the orbital path of Enceladus. Comparing our low-latitude profiles with this result, we see that water deposited in the lower atmosphere appears to be co-located with the edges of an

ionospheric peak extending from $1.52R_S$ to $1.62R_S$ in H_3^+ emission, within a couple of degrees of planetocentric latitude, as seen in Fig. 2.

Water ions flowing from the rings along magnetic field lines into the ionosphere cause the electron density to be reduced through rapid chemical recombination (quenching)⁷. Charged species derived from water also deplete H_3^+ because it protonates (charge-exchanges) quickly with molecules heavier than H and He (ref. 15); a drop in H_3^+ density (and thus intensity at latitudes where the most water is delivered to the planet) should therefore be visible. When mapping along field lines from the ionosphere to the equatorial plane, we found that each large trough corresponds to a major subdivision of the rings. In the same manner, the peaks in emission are found to map to prominent gaps in the rings. The reason for these peaks in relatively high intensity may be that these are regions in which the influx of water is severely reduced. If so, these peaks are not really peaks at all, but regions in which the ionosphere is quenched less than are the latitudes on either side.

The Cassini division, for example, which maps on average to about $2.1R_S$ in Fig. 2, occurs at latitudes where an increase in the emission of H_3^+ is clearly visible. Water influx from the A and B rings quenches the ionosphere at locations on either side of these latitudes, leading to the prominent peak seen in between. This occurs in both hemispheres, symmetrical about the magnetic equator of Saturn. One exception to this apparent correlation between water influx and H_3^+ emission is between the instability radii. Modelling suggests that water influx should peak at these regions¹¹, but our measurements clearly show these regions to lie within a peak in emission, rather than at the locations of quenching. A possible explanation for this anomaly is that

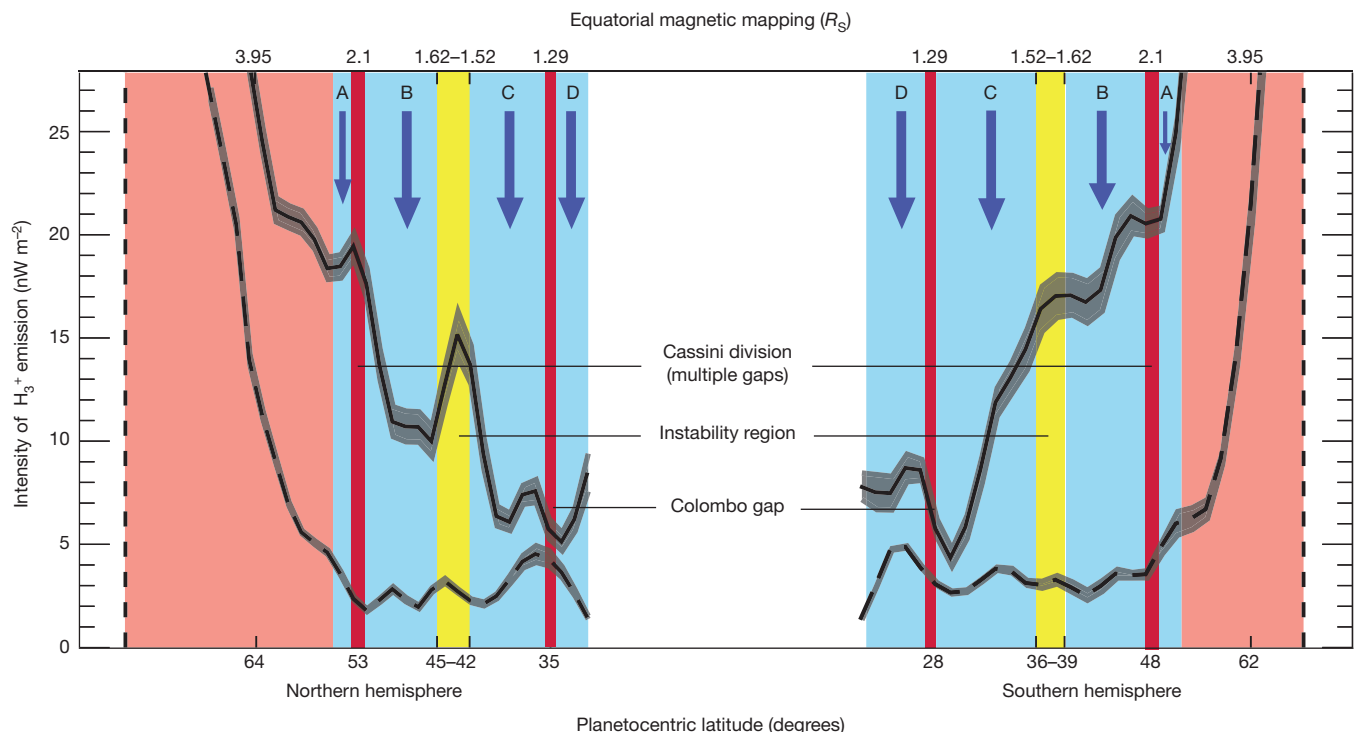


Figure 2 | Intensity of H_3^+ infrared emission as a function of position along Saturn's noon meridian. The horizontal axes show a scale of planetocentric latitude at the bottom and the planetocentric equatorial distances which those latitudes magnetically map to at the top. The y axis shows the intensity of H_3^+ emission of the two spectral lines that are shown, $\text{Q}(1,0^-)$ at $3.953\ \mu\text{m}$ (black line) and $\text{R}(2,2^-)$ at $3.622\ \mu\text{m}$ (dashed black line) with a central gap where the observed emission is swamped by solar photon reflection from the planetary rings. The 1-sigma errors in intensity measurements are denoted by the grey shading envelopes for each line. Latitude bands mapping along planetary magnetic field lines to the main ring subdivisions in the equatorial plane are shown blue (water influx), and the ring gaps are shown in red. The yellow

shading is the instability region between the stability limits, as discussed in the main text. High- to mid-latitude emission is shaded pink out to the limb of the planet (dashed vertical line), peaking in intensity at about $2\ \mu\text{W m}^{-2}$ for the $\text{Q}(1,0^-)$ line and at about $1\ \mu\text{W m}^{-2}$ for the $\text{R}(2,2^-)$ line (polar auroral/mid-latitude emission). The errors in latitude are on average 3° , mainly caused by the Earth's atmospheric attenuation (that is, seeing, of $0.4\ \text{s of arc}$). To remove additional errors, only the best seeing conditions were selected, such that the $\text{Q}(1,0^-)$ line derives from the co-addition of about 40% of the data set. However, owing to the weaker signal in the $\text{R}(2,2^-)$ line, about 90% of the data set was co-added, leading to a greater error in latitude but a reduction in intensity errors (see also Supplementary Information).

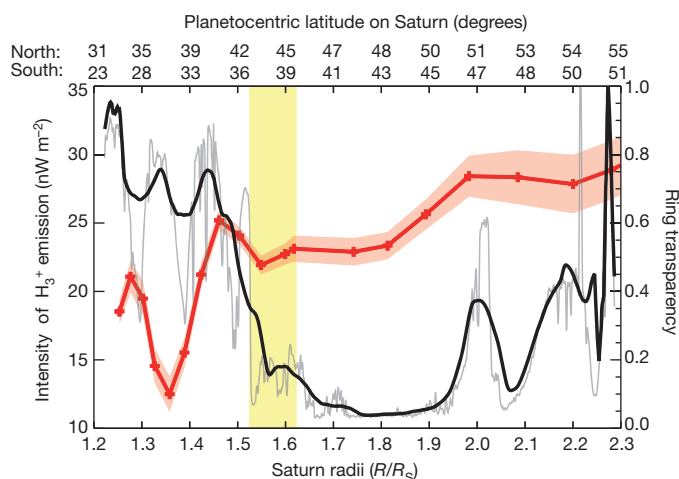


Figure 3 | Comparison of H_3^+ intensity and the transparency of Saturn's rings. The horizontal axes show latitude in each hemisphere at the top in degrees and the corresponding equatorial distance at the bottom, mapped along magnetic field lines. The H_3^+ intensities from the Q(1, 0[−]) and R(2, 2[−]) lines in both hemispheres are co-added to obtain a high signal-to-noise ratio, and are shown by the red line (pale red shading denotes the 1-sigma error in the intensity measurements); a line-of-sight correction has been applied. The black and grey lines show a smoothed and raw normalized photon transmission count (transparency), respectively, of the rings taken from Voyager 2 Ultraviolet Spectrometer stellar occultation archives¹⁸. The yellow shading indicates the instability region between the instability radii at 1.52 R_S and 1.62 R_S . The H_3^+ intensity shows peaks in regions of high ring transparency and troughs where transparency is low, indicating that H_3^+ emission is quenched when it is mapped to high-density regions of the rings. At higher latitudes, one pixel represents more degrees of latitude than at low latitudes, as depicted by the red crosses, which indicate the pixel spacing on the spectrograph charge-coupled device (CCD).

while significantly enhanced water influx occurs on the edges of this region, there could be low densities of water ion influx in between, leading to little or no reduction of H_3^+ density at latitudes mapping to it. The reduced water source here may be the result of the instability radii consuming the local supply at either side of the instability region, effectively cutting off the supply. However, as stated, the uncertainty in mapping here can affect the interpretation of this peak in emission, so we also explore other mechanisms that could create the observed features.

Another interpretation for the observations is that the peaks in intensity correspond to temperature increases in H_3^+ , whereas the troughs correspond to the natural background levels of H_3^+ emission produced by solar extreme-ultraviolet ionization. These temperature increases would be the result of Joule heating via the flow of charged particles, which requires that the resultant rise in H_3^+ intensity is large enough to overcome any quenching of H_3^+ density that may take place. Detailed modelling of the effects of ring rain are required to establish what the background H_3^+ emission intensity in these latitudes should be (on the basis of solar extreme-ultraviolet ionization alone), and whether or not the peaks in emission found here are equal to or higher than this level. The shadow cast by the rings is known to create variations in the ion density with latitude¹⁶, but it cannot explain the features seen here because the shadow falls behind the rings in ground-based geometry. As seen in Fig. 1, the reflection of sunlight by the rings obscures this region entirely.

Atmospheric models still continue to struggle to explain the electron density distribution^{6–8} and underestimate the observed temperature of the neutral gases in the upper atmosphere at low latitudes by many hundreds of Kelvin, in what is known colloquially as the ‘energy crisis’^{14,17}.

This shortfall highlights the acute need for a greater understanding of the systems acting upon the low-latitude atmosphere. The observations here are the first direct measure of Saturn's ionospheric reaction to ring-water input, rather than through measurements of the deeper atmosphere below it. Previously, only discrete features linked to water influx have been found, but here we can see that the scale of this interaction is global, and that water from the A, B and C rings can quench the ionosphere, mapping to over 30% of Saturn's surface area. The D-ring mapped latitudes are largely obscured, but their inclusion would imply that up to 43% of the planet is subjected to water influx.

Received 27 October 2012; accepted 22 February 2013.

1. Stallard, T. S. *et al.* Temperature changes and energy inputs in giant planet atmospheres: what we are learning from H_3^+ . *Phil. Trans. R. Soc.* **370**, 5213–5224 (2012).
2. Miller, S., Stallard, T., Melin, H. & Tennyson, J. H_3^+ cooling in planetary atmospheres. *Faraday Discuss.* **147**, 283–291 (2010).
3. Atreya, S. K., Donahue, T. M., Nagy, A. F., Waite, J. H. Jr & McConnell, J. C. Theory, measurements, and models of the upper atmosphere and ionosphere of Saturn. In *Saturn from Cassini-Huygens* (eds Dougherty, M., Esposito, L. & Krimigis, S.) 239–277 (Springer, 1984).
4. Smith, C. G. A., Aylward, A. D., Millward, G. H., Miller, S. & Moore, L. E. An unexpected cooling effect in Saturn's upper atmosphere. *Nature* **445**, 399–401 (2007).
5. Moore, L., Mueller-Wodarg, I., Galand, M., Kliore, A. & Mendillo, M. Latitudinal variations in Saturn's ionosphere: Cassini measurements and model comparisons. *J. Geophys. Res.* **115**, A11317, <http://dx.doi.org/10.1029/2010JA015692> (2010).
6. Connerney, J. E. P. & Waite, J. H. New model of Saturn's ionosphere with an influx of water from the rings. *Nature* **312**, 136–138 (1984).
7. Connerney, J. E. P. Magnetic connection for Saturn's rings and atmosphere. *Geophys. Res. Lett.* **13**, 773–776 (1986).
8. Wilson, G. R. & Waite, J. H. Jr. Kinetic modeling of the Saturn ring-ionosphere plasma environment. *J. Geophys. Res.* **94**, 17287–17298 (1989).
9. McLean, I. S. *et al.* in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 3354* (ed. Fowler, A. M.) 566–578 (SPIE, 1998).
10. Northrop, T. G. & Hill, J. R. Stability of negatively charged dust grains in Saturn's ring plane. *J. Geophys. Res.* **87**, 6045–6051 (1982).
11. Northrop, T. G. & Hill, J. R. The inner edge of Saturn's B ring. *J. Geophys. Res.* **88**, 6102–6108 (1983).
12. Burton, M. E., Dougherty, M. K. & Russell, C. T. Saturn's internal planetary magnetic field. *J. Geophys. Res.* **37**, 24105 (2010).
13. Luhmann, J. G., Johnson, R. E., Tokar, R. L., Ledvina, S. A. & Cravens, T. E. A model of the ionosphere of Saturn's rings and its implications. *Icarus* **181**, 465–474 (2006).
14. Coates, A. J. *et al.* Plasma electrons above Saturn's main rings: CAPS observations. *Geophys. Res. Lett.* **32**, L14S09, <http://dx.doi.org/10.1029/2005GL022694> (2005).
15. Miller, S. *et al.* H_3^+ : the driver of giant planet atmospheres. *Phil. Trans. R. Soc. Lond.* **364**, 3121–3137 (2006).
16. Mendillo, M. *et al.* Effects of ring shadowing on the detection of electrostatic discharges at Saturn. *Geophys. Res. Lett.* **32**, L14S09, <http://dx.doi.org/10.1029/2004GL021934> (2005).
17. Yelle, R. V. & Miller, S. in *Jupiter: the Planet, Satellites and Magnetosphere* (eds Bagenal, F., Dowling, T. & McKinnon, W.) 185–218 (Cambridge Univ. Press, 2004).
18. Lillie, C. F., Hord, C. W., Pang, K., Coffeen, D. L. & Hansen, J. E. The Voyager mission photopolarimeter experiment. *Space Sci. Rev.* **21**, 159–181 (1977).

Supplementary Information is available in the online version of the paper.

Acknowledgements The data presented here were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and NASA. The observations were made to support the Cassini auroral campaign. Ring profile data were provided by the Planetary Rings Node website¹⁸. Discussions within the international team led by T.S.S. on ‘Comparative Jovian Aeronomy’ have greatly benefited this work; this was hosted by the International Space Science Institute (ISSI). The UK Science and Technology Facilities Council (STFC) supported this work through the PhD Studentship of J.O'D. and grant support for T.S.S., H.M. and G.H.J.

Author Contributions J.O'D. analysed and interpreted the data and wrote the paper. T.S.S., S.M. and K.H.B. proposed and designed the study and collected the data. H.M. greatly aided the analysis and interpretation of data. S.W.H.C. provided the magnetic-mapping model and magnetospheric information. G.H.J. provided ring-plane information. J.S.D.B. provided context from Cassini VIMS observations. All authors assisted in the interpretation of data and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.O'D. (jod3@ion.le.ac.uk).

Photonic Floquet topological insulators

Mikael C. Rechtsman^{1*}, Julia M. Zeuner^{2*}, Yonatan Plotnik^{1*}, Yaakov Lumer¹, Daniel Podolsky¹, Felix Dreisow², Stefan Nolte², Mordechai Segev¹ & Alexander Szameit²

Topological insulators are a new phase of matter¹, with the striking property that conduction of electrons occurs only on their surfaces^{1–3}. In two dimensions, electrons on the surface of a topological insulator are not scattered despite defects and disorder, providing robustness akin to that of superconductors. Topological insulators are predicted to have wide-ranging applications in fault-tolerant quantum computing and spintronics. Substantial effort has been directed towards realizing topological insulators for electromagnetic waves^{4–13}. One-dimensional systems with topological edge states have been demonstrated, but these states are zero-dimensional and therefore exhibit no transport properties^{11,12,14}. Topological protection of microwaves has been observed using a mechanism similar to the quantum Hall effect¹⁵, by placing a gyromagnetic photonic crystal in an external magnetic field⁵. But because magnetic effects are very weak at optical frequencies, realizing photonic topological insulators with scatter-free edge states requires a fundamentally different mechanism—one that is free of magnetic fields. A number of proposals for photonic topological transport have been put forward recently^{6–10}. One suggested temporal modulation of a photonic crystal, thus breaking time-reversal symmetry and inducing one-way edge states¹⁰. This is in the spirit of the proposed Floquet topological insulators^{16–19}, in which temporal variations in solid-state systems induce topological edge states. Here we propose and experimentally demonstrate a photonic topological insulator free of external fields and with scatter-free edge transport—a photonic lattice exhibiting topologically protected transport of visible light on the lattice edges. Our system is composed of an array of evanescently coupled helical waveguides²⁰ arranged in a graphene-like honeycomb lattice^{21–26}. Paraxial diffraction of light is described by a Schrödinger equation where the propagation coordinate (z) acts as ‘time’²⁷. Thus the helicity of the waveguides breaks z -reversal symmetry as proposed for Floquet topological insulators. This structure results in one-way edge states that are topologically protected from scattering.

Paraxial propagation of light in photonic lattices is described by the Schrödinger-type equation:

$$i\partial_z \psi(x, y, z) = -\frac{1}{2k_0} \nabla^2 \psi(x, y, z) - \frac{k_0 \Delta n(x, y, z)}{n_0} \psi(x, y, z) \quad (1)$$

where $\psi(x, y, z)$ is the electric field envelope function defined by $E(x, y, z) = \psi(x, y, z) \exp(ik_0 z - i\omega t)$; E is the electric field, \mathbf{x} is a unit vector and t is time; the Laplacian, ∇^2 , is restricted to the transverse (x – y) plane; $k_0 = 2\pi n_0/\lambda$ is the wavenumber in the ambient medium; $\omega = 2\pi c/\lambda$ is the optical frequency; and c and λ are respectively the velocity and wavelength of light. Our ambient medium is fused silica with refractive index $n_0 = 1.45$, and $\Delta n(x, y, z)$ is the ‘effective potential’, that is, the deviation from the ambient refractive index. The array is fabricated using the femtosecond laser writing method; each elliptical waveguide has a cross-section with major and minor axis diameters of 11 μm and 4 μm , respectively. The photonic lattice is an array of evanescently-coupled waveguides arranged in a honeycomb structure with nearest-neighbour spacing of 15 μm . The total propagation length (in the z direction) is 10 cm, which corresponds to the wavefunction ψ

of a single waveguide mode completing ~ 20 cycles in phase while propagating from $z = 0$ to $z = 10$ cm. The increase in refractive index associated with the waveguides is $\Delta n = 7 \times 10^{-4}$. The quantum mechanical analogue of equation (1) describes the propagation of a quantum particle that evolves in time—for example, an electron in a solid. The waveguides act like potential wells, similarly to nuclei of atoms in solids. Thus, the propagation of light in the array of helical waveguides as it propagates in the z direction is equivalent to the temporal evolution of an electron as it moves through a two-dimensional lattice with atoms that are rotating in time.

A microscope image of the input facet of the photonic lattice is shown in Fig. 1a, and a diagram of the helical waveguides arranged in a honeycomb lattice is shown in Fig. 1b. The period (or pitch) of the helical waveguides is sufficiently small that a guided mode is adiabatically drawn along with a waveguide as it curves. We therefore transform the coordinates into a reference frame where the waveguides are invariant in the z direction (i.e., straight), namely: $x' = x + R \cos(\Omega z)$, $y' = y + R \sin(\Omega z)$ and $z' = z$, where R is the helix radius and $\Omega = 2\pi/Z$ = $2\pi/1$ cm is the frequency of rotation ($Z = 1$ cm being the period). In the transformed coordinates, the light evolution is described by:

$$i\partial_{z'} \psi' = -\frac{1}{2k_0} (\nabla' + i\mathbf{A}(z'))^2 \psi' - \frac{k_0 R^2 \Omega^2}{2} \psi' - \frac{k_0 \Delta n(x', y')}{n_0} \psi' \quad (2)$$

where $\psi' = \psi(x', y', z')$, and $\mathbf{A}(z') = k_0 R \Omega [\sin(\Omega z'), -\cos(\Omega z'), 0]$ is equivalent to a vector potential associated with a spatially homogeneous electric field of circular polarization. The adiabaticity of the guided modes and the presence of the vector potential yield a coupled mode (tight-binding) equation, via the Peierls substitution¹⁰:

$$i\partial_{z'} \psi_n(z') = \sum_{\langle m \rangle} c e^{i\mathbf{A}(z') \cdot \mathbf{r}_{mn}} \psi_m(z') \quad (3)$$

where the summation is taken over neighbouring waveguides; $\psi_n(z')$ is the amplitude in the n th waveguide, c is the coupling constant between waveguides and \mathbf{r}_{mn} is the displacement between waveguides m and n . Because the right-hand side of equation (3) is z -dependent, there are no static eigenmodes. Rather, the solutions are described using Floquet modes, of the form $\psi_n(z') = \exp(i\beta z') \varphi_n(z')$, where the function $\varphi_n(z')$ is Z -periodic¹⁸. This yields the spectrum of β (the Floquet eigenvalues or ‘quasi-energies’) as a function of the Bloch wavevector, (k_x, k_y) , as well as their associated Floquet eigenmodes. Floquet eigenmodes in the z direction are equivalent to Bloch modes in the x – y plane. Therefore, our input beam (initial wavefunction) excites a superposition of Floquet modes whose population does not change over the course of propagation^{17,18}. The band structure for the case of non-helical waveguides ($R = 0$) is shown in Fig. 1c. The conical intersections between the first and second bands are the ‘Dirac points’²⁸, a feature of graphene that makes it semi-metallic. When the waveguides are made helical ($R > 0$), a bandgap in the Floquet spectrum opens, as shown in Fig. 1d, and the photonic lattice becomes analogous to an insulator—indeed, to a Floquet topological insulator. As we show below, this structure possesses topologically protected edge states.

¹Department of Physics and the Solid State Institute, Technion – Israel Institute of Technology, Haifa 32000, Israel. ²Institute of Applied Physics, Abbe Center of Photonics, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany.

*These authors contributed equally to this work.

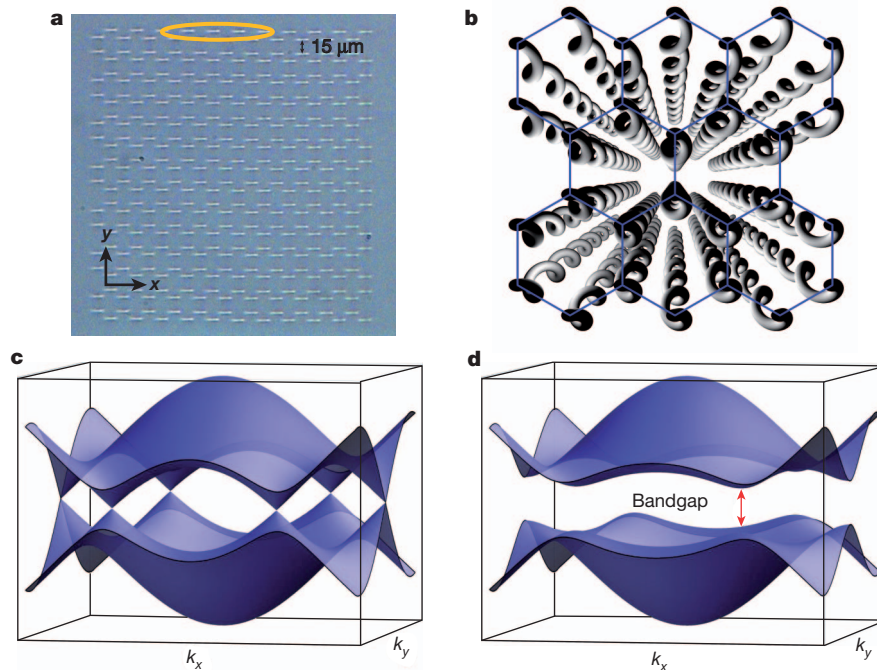


Figure 1 | Geometry and band structure of honeycomb photonic Floquet topological insulator lattice. **a**, Microscope image of the input facet of the photonic lattice, showing honeycomb geometry with 'zig-zag' edge terminations on the top and bottom, and 'armchair' terminations on the left and right sides. Scale bar at top right, 15 μm . The yellow ellipse indicates the position and shape of the input beam to this lattice. **b**, Sketch of the helical waveguides. Their rotation axis is in the z direction, with radius R and period

Z . **c**, Band structure (β versus (k_x, k_y)) for the case of non-helical waveguides comprising a honeycomb lattice ($R = 0$). Note the band crossings at the Dirac point. **d**, Bulk band structure for the photonic topological insulator: helical waveguides with $R = 8 \mu\text{m}$ arranged in a honeycomb lattice. Note the bandgap opening up at the Dirac points (labelled with the red, double-ended arrow), which corresponds to the bandgap in a Floquet topological insulator.

We calculate the edge band structure by using a unit cell that is periodic in the x direction but finite in the y direction, ending with two 'zig-zag' edges (infinite in the x direction). The zig-zag edge is one of three typical edge terminations of the honeycomb lattice; the other two are the 'armchair edge' and the 'bearded edge'. Note that the presence of chiral edge states can be derived using the bulk-edge correspondence principle by calculating the Chern number^{4,5,17,29}. In our sample (see Fig. 1a), the top and bottom edges are zig-zag edges and the right and left edges are armchair edges. The band structure of the zig-zag edge is presented in Fig. 2a for the case where the waveguides are not helical ($R = 0$). There are two sets of states, one per edge. Their dispersion curves are flat and completely coincide (that is, they are degenerate with one another), residing between $k_x = 2\pi/3a$ and $k_x = 4\pi/3a$, occupying one-third of k_x space, where $a = 15\sqrt{3} \mu\text{m}$ is the lattice constant. The Floquet band structure when the lattice is helical with $R = 8 \mu\text{m}$ is shown in Fig. 2b. Here, the edge states are no longer degenerate, but now have opposite slopes. Specifically, the transverse group velocity

(i.e., the group velocity in the $(x-y)$ plane) on the top edge is now directed to the right, and on the bottom edge to the left, corresponding to clockwise circulations. However, there are no edge states whatsoever circulating in the anti-clockwise direction. Hence, the edge states presented in Fig. 2b are the topologically protected edge states of a Floquet topological insulator. The lack of a counter-propagating edge state on a given edge directly implies that any edge-defect (or disorder) cannot backscatter, as there is no backwards-propagating state available into which to scatter, contrary to the case of $R = 0$, where there are multiple states into which scattering is possible. This is the essence of why topological protection occurs. The transverse group velocity (for brevity, we henceforth drop 'transverse') of these edge states has a non-trivial dependence on the helix radius, R . For small R , the group velocity of the edge states increases, but eventually it reaches a maximum and decreases again. Before the group velocity crosses zero, the Chern number is calculated to be -1 (indicating the presence of a clockwise edge state, as seen in Fig. 2b). However, after the group velocity crosses

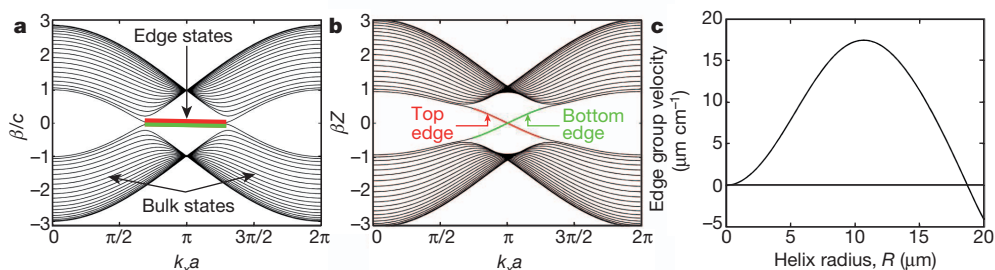


Figure 2 | Dispersion curves of the edge states, highlighting the unique dispersion properties of the topologically protected edge states for helical waveguides in the honeycomb lattice. **a**, Band structure of the edge states on the top and bottom of the array when the waveguides are straight ($R = 0$). The dispersion of both top and bottom edge states (red and green curves) is flat, therefore they have zero group velocity. The bands of the bulk honeycomb lattice

are drawn in black. **b**, Dispersion curves of the edge states in the Floquet topological insulator for helical waveguides with $R = 8 \mu\text{m}$: the band gap is open and the edge states acquire non-zero group velocity. These edge states reside strictly within the bulk band gap of the bulk lattice (drawn in black). **c**, Group velocity (slope of green and red curves) versus helix radius, R , of the helical waveguides comprising the honeycomb lattice. The maximum occurs at $R = 10.3 \mu\text{m}$.

zero—at which point the band gap closes—the Chern number is 2 (indicating the presence of two anti-clockwise edge states, as confirmed by calculations). The R dependence of the group velocity is shown in Fig. 2c, where we plot the group velocity of the topologically protected edge state at $k_x = \pi/a$ versus R . The maximum calculated group velocity is at $R = 10.3 \mu\text{m}$.

To demonstrate these edge states experimentally, we launch a beam with an elliptic profile of wavelength 633 nm such that it is incident on the top row of waveguides in an array with helix radius $R = 8 \mu\text{m}$. The position of the input beam is indicated by the ellipse in Fig. 1a. The light distribution emerging from the output facet is presented in Fig. 3a–d, with the shape and position of the input beam indicated by a yellow ellipse. In Fig. 3a, the beam emerges at the upper-right corner of the lattice, having moved along the upper edge. When we move the position of the input beam horizontally to the right, the output beam moves down along the vertical right edge, as shown in Fig. 3b. The beam emerging from the lattice remains confined to the edge, not spreading into the bulk and without any backscattering. Moving the position of the input beam further rightward makes the output beam move farther down along the side edge, as shown in Fig. 3c and d. Clearly, the input beam has moved along the top edge, encountered the corner, and then continued moving downward along the right edge. We show this behaviour in beam-propagation-method (BPM) simulations³⁰, solving equation (1) (see Supplementary Video 1). The central observation of these experimental results is that the corner (which is in essence a strong defect) does not backscatter light. Indeed, no optical intensity is evident along the top edge at the output facet, after having backscattered from the corner. Furthermore, no scattering into the bulk of the array is observed (owing to the presence of a bulk band-gap). These observations provide strong evidence of topological protection of the edge state.

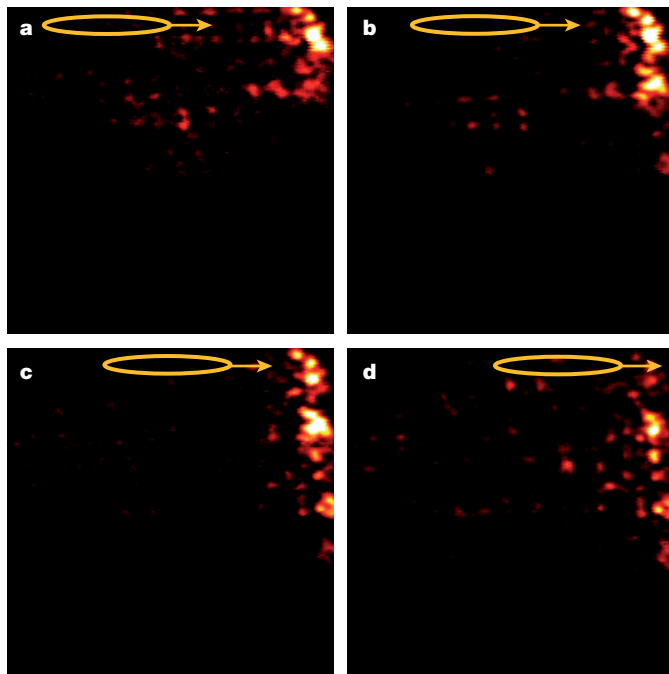


Figure 3 | Light emerging from the output facet of the waveguide array as the input beam is moved rightwards, along the top edge of the waveguide array. The yellow ellipse at the top of each panel shows the position of the input beam (which is at the top of the array, see Fig. 1a), which is moved progressively to the right in a–d. The beam propagates along the top edge of the array (which is in the zig-zag configuration), hits the corner, and clearly moves down the vertical edge (which is in the armchair configuration). Note that the wavepacket shows no evidence of backscattering or bulk scattering due to its impact with the corner of the lattice. This scattering of the edge state is prevented by topological protection.

Further evidence follows from the fact that light stays confined to the side edge of the array as it propagates downwards. This edge is in the armchair geometry, which, for straight waveguides ($R = 0$) does not allow edge confinement at all (that is, no edge states). However, when $R > 0$, edge state dispersion calculations reveal that a confined edge state emerges. This is essential for the topological protection because it prevents transport into the bulk of the lattice.

We now experimentally examine the behaviour of the topological edge states as the helix radius, R , is varied. We find that the group velocity reaches a maximum and then returns to zero as R is increased, in accordance with Fig. 2c. To investigate this, we fabricate a series of honeycomb lattices of helical waveguides with increasing values of R , cut in a triangular shape (Fig. 4a). We first examine light propagation in the lattice with non-helical waveguides (that is, $R = 0$; Fig. 4b). Launching a beam into the waveguide at the upper-left corner of the triangle (circled) excites two types of eigenstates: (1) bulk states extending to the corner, and (2) edge states that meet at the corner. As the light propagates in the array, the excited bulk states lead to some degree of spreading into the bulk (the excitation of these bulk modes can be eliminated by engineering the beam to only overlap with eigenstates confined to the edge). In contrast, the edge states do not spread into the bulk, and, because the edge states are all degenerate (Fig. 2a), they do not cause spreading along the edges either (that is, zero group velocity). Figure 4b shows the intensity at the output facet highlighting this effect: while some light has diffracted into the bulk, the majority remains at the corner waveguide. This is also shown in simulations (where the animation evolves by sweeping through the z coordinate from $z = 0 \text{ cm}$ to $z = 10 \text{ cm}$); see Supplementary Video 2.

When the helical waveguides have clockwise rotation, the edge states are no longer degenerate. In fact, the lattice now has a set of edge states that propagate only clockwise on the circumference of the triangle. Light at the corner no longer remains there, and moves along the edge. Figure 4b–j shows the output facet of the lattice for increasing radius R . For $R = 8 \mu\text{m}$, the wave packet wraps around the corner of the triangle and moves along the opposite edge (Fig. 4f) (the corresponding simulation is shown in Supplementary Video 3; the loss of intensity over the course of propagation is due to bending/radiation losses). Importantly, the light is not backscattered even when it hits the acute corner, owing to the lack of a counter-propagating edge state. This is a key example of topological protection against scattering. For $R = 12 \mu\text{m}$, the wavepacket moves along the edge, but with a slower group velocity. This is consistent with the prediction that the group velocity of the edge state reaches a maximum at $R = 10.3 \mu\text{m}$ and thereafter decreases with increasing radius. The experiments suggest that the maximal group velocity is achieved between $6 \mu\text{m}$ and $10 \mu\text{m}$, while the theoretical result ($10.3 \mu\text{m}$) is well within experimental error, given that this is a prediction from coupled-mode theory. Exact simulations confirm the experimental result.

By $R = 16 \mu\text{m}$, bending losses are large, leading to leakage of optical power into scattering modes (accounting for the large background signal). The bending losses for $R = 4 \mu\text{m}$, $8 \mu\text{m}$, $12 \mu\text{m}$ and $16 \mu\text{m}$ were found to be, respectively, 0.03 dB cm^{-1} , 0.5 dB cm^{-1} , 1.7 dB cm^{-1} and 3 dB cm^{-1} . Recall that each lattice has propagation length $z = 10 \text{ cm}$. The large background signal prevents us from experimenting with larger R , where we would expect two anti-clockwise-propagating edge states, as discussed earlier. As shown in Fig. 4j, the group velocity of the wavepacket approaches zero and therefore the optical power remains at the corner waveguide. These observations clearly demonstrate the presence of one-way edge states on the boundary of the photonic lattice that behave according to theory. Note that for different initial beams—the elliptical beam of Fig. 3, and the single-waveguide excitation of Fig. 4—the topological edge state behaves exactly as the model predicts, providing experimental proof of the existence of the topological edge state.

To demonstrate the z dependence of the wavepacket as it propagates along the edge, we turn to a combination of experimental results and

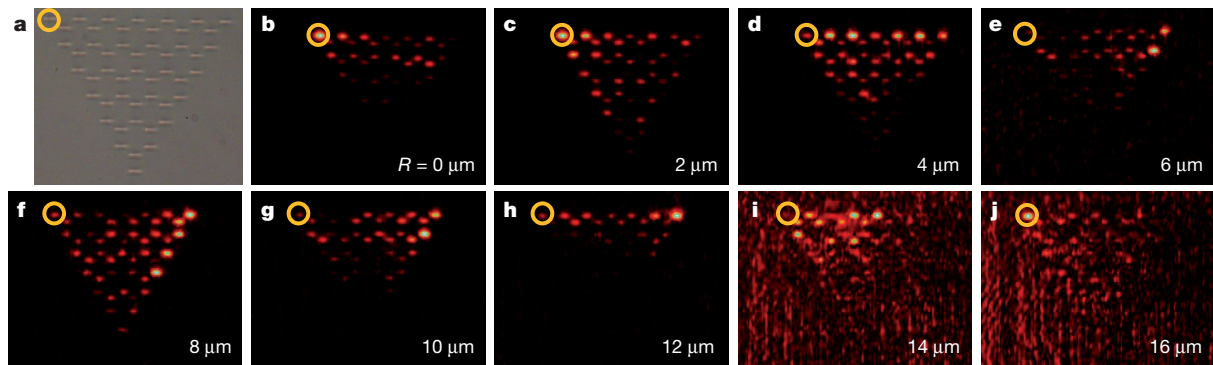


Figure 4 | Experiments highlighting light circulation along the edges of a triangular-shaped lattice of helical waveguides arranged in a honeycomb geometry. **a**, Microscope image of the honeycomb lattice in the triangular configuration. **b–j**, Light emerging from the output facet of the photonic lattice (after 10 cm of propagation) for increasing helix radius, R (given at bottom right of each panel), at wavelength 633 nm. The light is initially launched into the waveguide at the upper-left corner (on the input facet of the array, indicated by a yellow circle). At $R = 0$ (**b**), the initial beam excites a confined defect mode at

the corner. As the radius is increased (**c–j**), light is moving along the edge by virtue of a topological edge mode. It reaches its maximum displacement near $R = 8 \mu\text{m}$ (**f**). The light wraps around the corner of the triangle and is not backscattered at all: this is a clear example of topological protection against scattering. As R is increased further, the light exhibits a decreasing group velocity as a function of R , and finally stops near $R = 16 \mu\text{m}$. The large degree of background noise in **i** and **j** is due to high bending losses of the waveguides as a result of coupling to free-space scattering modes.

simulations of equation (1)³⁰. We examine a lattice with a defect on the edge in the form of a ‘missing’ waveguide (Fig. 5a). Because of topological protection, the wavepacket should simply propagate around the missing waveguide (the defect) without backscattering. An experimental image of the output facet is shown in Fig. 5b (for $R = 8 \mu\text{m}$). The excited waveguide is at the top right, and the edge state propagates clockwise, avoiding the defect, and eventually hitting the next corner. In Fig. 5c–h we show simulations for the optical intensity at $z = 0, 2, 4, 6, 8, 10$ cm, respectively. The wavepacket clearly propagates around the

defect, continuing forward without backscattering. Note that the simulated wavepacket has progressed slightly farther than that in the experiment. This is a result of small uncertainty in the coupling constant, c . Taken together, these data show the progression of topologically protected modes as they travel along the edge.

Photonic Floquet topological insulators have the potential to provide an entirely new platform for probing and understanding topological protection. For example, our photonic lattices have the same geometry as photonic crystal fibres, and thus these systems are likely to exhibit robust topologically protected states. Many interesting open questions are prompted, concerning (for example) the behaviour of entangled photons in a topologically protected system, the effect of interactions on the non-scattering behaviour, or the possibility of simulating photonic Majorana fermions for applications in robust quantum computing. The realization of a photonic Floquet topological insulator in our relatively simple classical system will enable these questions, as well as many others, to be addressed.

Received 17 December 2012; accepted 12 March 2013.

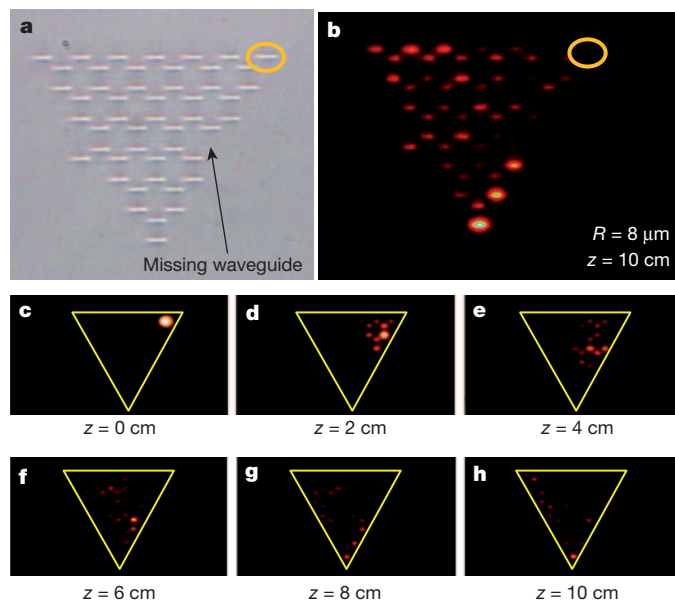


Figure 5 | Experiments and simulations showing topological protection in the presence of a defect. The lattice is triangular-shaped, and the waveguides are helical with $R = 8 \mu\text{m}$. **a**, Microscope image of photonic lattice with a missing waveguide (acting as a defect, arrowed) on the rightmost zig-zag edge. A light beam of $\lambda = 633$ nm is launched into the single waveguide at the upper-right corner (on the far side of the array, surrounded by a yellow circle). **b**, Experimental image of light emerging from the output facet after $z = 10$ cm of propagation, showing no backscattering despite the presence of the defect (a signature of topological protection). **c–h**, Numerical simulations of light propagation through the lattice at various propagation distances (respectively $z = 0$ cm, 2 cm, 4 cm, 6 cm, 8 cm and 10 cm). After minimal bulk scattering, the light propagates along the edge, encounters the defect, propagates around it, and continues past it without scattering, in agreement with **b**.

1. Kane, C. L. & Mele, E. J. Quantum spin Hall effect in graphene. *Phys. Rev. Lett.* **95**, 226801 (2005).
2. König, M. *et al.* Quantum spin Hall insulator state in HgTe quantum wells. *Science* **318**, 766–770 (2007).
3. Hsieh, D. *et al.* A topological Dirac insulator in a quantum spin Hall phase. *Nature* **452**, 970–974 (2008).
4. Haldane, F. D. M. & Raghu, S. Possible realization of directional optical waveguides in photonic crystals with broken time-reversal symmetry. *Phys. Rev. Lett.* **100**, 013904 (2008).
5. Wang, Z., Chong, Y., Joannopoulos, J. D. & Soljacic, M. Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature* **461**, 772–775 (2009).
6. Koch, J., Houck, A. A., Hur, K. L. & Girvin, S. M. Time-reversal-symmetry breaking in circuit-QED-based photon lattices. *Phys. Rev. A* **82**, 043811 (2010).
7. Umucalilar, R. O. & Carusotto, I. Artificial gauge field for photons in coupled cavity arrays. *Phys. Rev. A* **84**, 043804 (2011).
8. Hafezi, M., Demler, E. A., Lukin, M. D. & Taylor, J. M. Robust optical delay lines with topological protection. *Nature Phys.* **7**, 907–912 (2011).
9. Khanikaev, A. B. *et al.* Photonic topological insulators. *Nature Mater.* **12**, 233–239 (2012).
10. Fang, K., Yu, Z. & Fan, S. Realizing effective magnetic field for photons by controlling the phase of dynamic modulation. *Nature Photon.* **6**, 782–787 (2012).
11. Kraus, Y. E., Lahini, Y., Ringel, Z., Verbin, M. & Zilberberg, O. Topological states and adiabatic pumping in quasicrystals. *Phys. Rev. Lett.* **109**, 106402 (2012).
12. Kitagawa, T. *et al.* Observation of topologically protected bound states in photonic quantum walks. *Nature Commun.* **3**, 882 (2012).
13. Lu, L., Joannopoulos, J. D. & Soljacic, M. Waveguiding at the edge of a three-dimensional photonic crystal. *Phys. Rev. Lett.* **108**, 243901 (2012).
14. Malkova, N., Hromada, I., Wang, X., Bryant, G. & Chen, Z. Observation of optical Shockley-like surface states in photonic superlattices. *Opt. Lett.* **34**, 1633–1635 (2009).

15. Klitzing, K. v., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
16. Oka, T. & Aoki, H. Photovoltaic Hall effect in graphene. *Phys. Rev. B* **79**, 081406 (2009).
17. Kitagawa, T., Berg, E., Rudner, M. & Demler, E. Topological characterization of periodically driven quantum systems. *Phys. Rev. B* **82**, 235114 (2010).
18. Lindner, N. H., Refael, G. & Galitski, V. Floquet topological insulator in semiconductor quantum wells. *Nature Phys.* **7**, 490–495 (2011).
19. Gu, Z., Fertig, H. A., Arovas, D. P. & Auerbach, A. Floquet spectrum and transport through an irradiated graphene ribbon. *Phys. Rev. Lett.* **107**, 216601 (2011).
20. Szameit, A. & Nolte, S. Discrete optics in femtosecond-laser-written photonic structures. *J. Phys. B* **43**, 163001 (2010).
21. Peleg, O. *et al.* Conical diffraction and gap solitons in honeycomb photonic lattices. *Phys. Rev. Lett.* **98**, 103901 (2007).
22. Bahat-Treidel, O., Peleg, O. & Segev, M. Symmetry breaking in honeycomb photonic lattices. *Opt. Lett.* **33**, 2251–2253 (2008).
23. Ablowitz, M. J., Nixon, S. D. & Zhu, Y. Conical diffraction in honeycomb lattices. *Phys. Rev. A* **79**, 053830 (2009).
24. Fefferman, C. L. & Weinstein, M. I. Honeycomb lattice potentials and Dirac points. *J. Am. Math. Soc.* **25**, 1169–1220 (2012).
25. Rechtsman, M. C. *et al.* Strain-induced pseudomagnetic field and photonic Landau levels in dielectric structures. *Nature Photon.* **7**, 153–158 (2013).
26. Crespi, A., Corrielli, G., Della Valle, G., Osellame, R. & Longhi, S. Dynamic band collapse in photonic graphene. *New J. Phys.* **15**, 013012 (2013).
27. Lederer, F. *et al.* Discrete solitons in optics. *Phys. Rep.* **463**, 1–126 (2008).
28. Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
29. Zak, J. Berry's phase for energy bands in solids. *Phys. Rev. Lett.* **62**, 2747–2750 (1989).
30. Kawano, K. & Kitoh, T. *Introduction to Optical Waveguide Analysis: Solving Maxwell's Equation and the Schrödinger Equation* (Wiley & Sons, 2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements M.C.R. is grateful to the Azrieli Foundation for the Azrieli fellowship while at the Technion. M.S. acknowledges the support of the Israel Science Foundation, the USA-Israel Binational Science Foundation, and an Advanced Grant from the European Research Council. A.S. thanks the German Ministry of Education and Research (Center for Innovation Competence program, grant 03Z1HN31) and the Thuringian Ministry for Education, Science and Culture (Research group Spacetime, grant 11027-514) for support. The authors thank S. Raghu and T. Pereg-Barnea for discussions.

Author Contributions The idea was conceived by Y.P. and M.C.R. The theory was investigated by M.C.R. and Y.P. The fabrication was carried out by J.M.Z. The experiments were carried out by M.C.R., Y.P. and J.M.Z. All authors contributed considerably.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C.R. (mcrworld@gmail.com) and Y.P. (yonatanplotnik@gmail.com).

Recent temperature extremes at high northern latitudes unprecedented in the past 600 years

Martin P. Tingley¹ & Peter Huybers¹

Recently observed extreme temperatures at high northern latitudes^{1–7} are rare by definition, making the longer time span afforded by climate proxies important for assessing how the frequency of such extremes may be changing. Previous reconstructions of past temperature variability have demonstrated that recent warmth is anomalous relative to preceding centuries^{2,8,9} or millennia¹⁰, but extreme events can be more thoroughly evaluated using a spatially resolved approach that provides an ensemble of possible temperature histories^{11,12}. Here, using a hierarchical Bayesian analysis^{13,14} of instrumental, tree-ring, ice-core and lake-sediment records, we show that the magnitude and frequency of recent warm temperature extremes at high northern latitudes are unprecedented in the past 600 years. The summers of 2005, 2007, 2010 and 2011 were warmer than those of all prior years back to 1400 (probability $P > 0.95$), in terms of the spatial average. The summer of 2010 was the warmest in the previous 600 years in western Russia ($P > 0.99$) and probably the warmest in western Greenland and the Canadian Arctic as well ($P > 0.90$). These and other recent extremes greatly exceed those expected from a stationary climate, but can be understood as resulting from constant space–time variability about an increased mean temperature.

Exceptionally high temperatures have recently been observed in Russia^{1–4}, Greenland^{5,6} and other locations at high northern latitudes⁷. Palaeoclimate reconstructions demonstrate that these temperatures are anomalous^{2,8–10}, but ascertaining whether they are unprecedented requires a treatment beyond that supported by standard measures of uncertainty^{11,12}. At issue is how warm it could have been at each location and each year in the past, given the uncertainty in each temperature estimate. For example, a temperature $T_0 = 33^\circ\text{C}$ is almost certainly higher than the i th realization of $T_i = 30 + \varepsilon_i$, where ε is a random variable drawn from a normal distribution with zero mean and unit variance, but it is unlikely to be higher than the maximum of 1,000 independent realizations of T_i .

The probability that a given interval contains the most extreme temperature in the span of a reconstruction is generally quantified using ensemble-based reconstruction methods because they can be used to estimate simultaneous, or pathwise, uncertainty intervals^{11,15}, and provide direct probabilistic assessments of extremes¹². Ensemble-based reconstructions have recently been used to evaluate extremes in spatially averaged temperature using bootstrap methods¹⁶, Bayesian principal component regression¹⁷ and realizations drawn from global climate model simulations^{18,19}.

Here we use a hierarchical Bayesian model^{13,14,20–22} to perform a $5^\circ \times 5^\circ$ spatial resolution analysis of the past 600 years of summer temperatures at high northern latitudes (see Methods). This approach gives a spatially resolved ensemble of temperature reconstructions, each of which is equally likely given the available instrumental and proxy data (see Methods). This approach also provides estimates of temperature and its uncertainty at locations without instrumental or proxy observations. We can therefore assign probabilities to years being warmest or coldest at each spatial location^{11,12} and assess recent extremes in the context of space–time climate variability.

The reconstruction is based on April–September averages of the Climate Research Unit's gridded instrumental data set²³ and 125 annually resolved proxy time series: 18 ice-core $\delta^{18}\text{O}$ series, 11 log-transformed annual varve thickness series from lake-floor sediment cores and 96 gridded maximum latewood tree-ring density series. We infer April–September temperature anomalies from 1400 to 2011 at all grid boxes between 45°N and 85°N that contain some fraction of land. The requirement that the proxy series be annually resolved, coupled with the need for sufficient coverage to produce a spatially complete reconstruction, necessitates that we limit the study to the past 600 years.

For temperature anomalies averaged over land areas north of 45°N , the years 2005, 2007, 2010 and 2011 are each warmer than all other years in the 1400–2004 interval ($P > 0.95$; Fig. 1a), whereas on decadal timescales, intervals subsequent to 1996 are warmer than all previous intervals ($P > 0.95$; Fig. 1b). These results support and extend previous findings regarding unprecedented warmth^{2,9,10} by explicitly considering whether a given anomaly is higher than those for all other years. Spatially averaged temperature estimates become more uncertain at earlier times, particularly before 1850 when the longest instrumental records drop out of the analysis (Fig. 1a and b). The unweighted average across the available instrumental observations closely agrees with the spatially averaged estimates from all data types back to about 1900, but not before. As the early instrumental observations are sparsely and heterogeneously distributed in space, the spatially complete reconstruction afforded by the statistical model (see Methods) and the proxy indicators of temperature become increasingly important towards earlier times.

Centennial-timescale variations can be characterized by linear trends (Fig. 1c). The linear rate of warming in the spatial-mean time series over the interval 1912–2011 is 1.06°C per century (90% uncertainty is 1.03°C to 1.08°C). With $P > 0.99$, the largest 100-year warming trend is centred after 1900, indicating that recent rates of warming are unprecedented in the past 600 years. The 100-year cooling trend with the largest magnitude occurred between 1547 and 1646 at a rate of -0.71°C per century (90% uncertainty is -0.88°C to -0.54°C). Interestingly, the warming trend has only recently exceeded the magnitude of this earlier cooling trend, such that if the reconstruction is curtailed at 2001, the cooling has a larger magnitude rate of change in 49% of the ensemble members. This early cooling trend, whose magnitude rivals that of the recent warming, also indicates that the proxy record is capable of capturing large, persistent temperature variations.

To investigate the distribution of extremes in both space and time, we calculate the probability that each year was warmest or coldest for each spatial location (Fig. 2). Whereas the 2000s had a high probability of containing the warmest year in many locations throughout northeastern Canada, Greenland, eastern Russian and the land areas surrounding the Bering Sea, the 1990s contained the warmest year predominantly in northwestern North America, and, to a lesser extent, central Russia (Fig. 2a). Averaging across the ensemble, we find that the warmest year occurs in 1990 or later for 60% of all locations, and in 2000 or later for 45%, demonstrating that the recent extreme warmth is widespread (Fig. 2b). There is no such clear trend in the occurrences of

¹Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA.

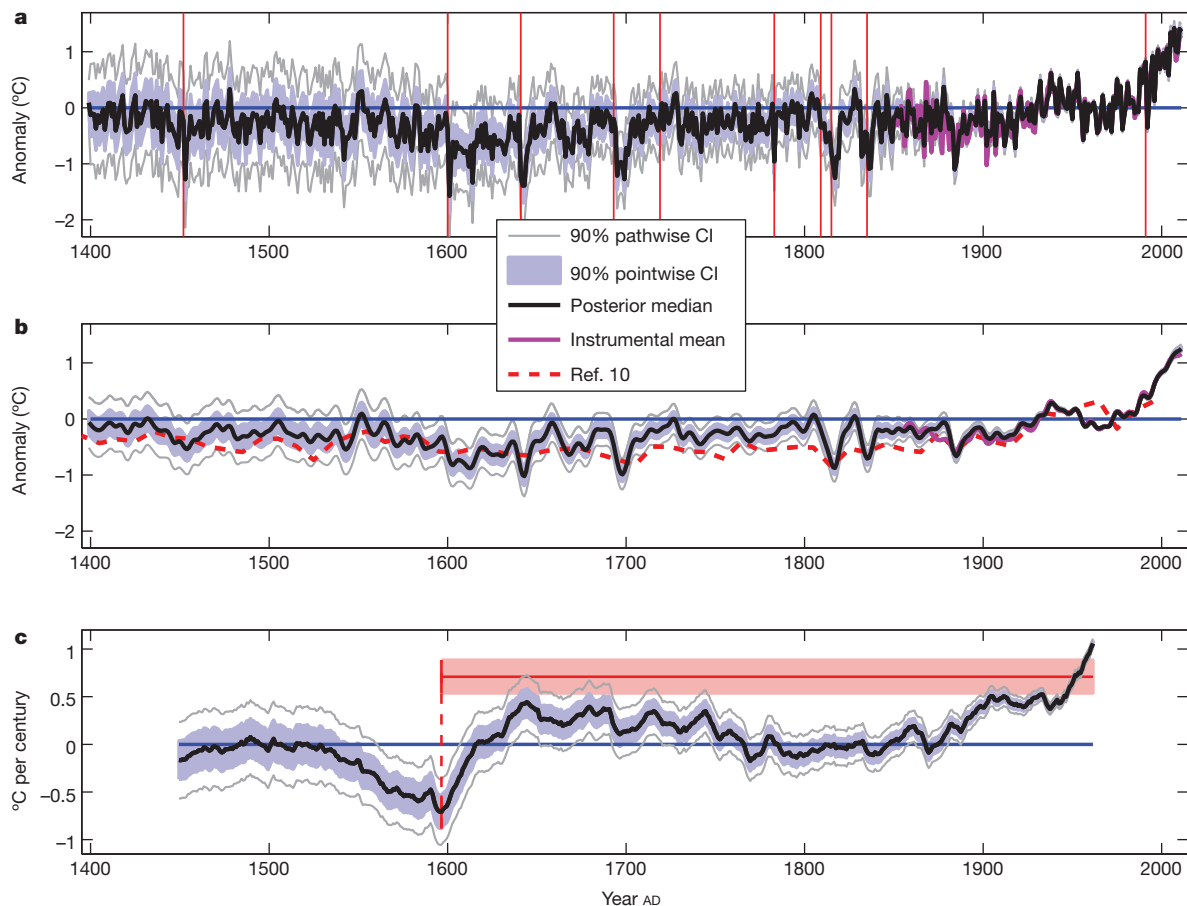


Figure 1 | Time series of temperature anomalies and centennial slopes.

a, Average land temperature between 45° N and 85° N (black), 90% pointwise (blue shading) and pathwise (grey) credible intervals²⁰ (see Methods); the unweighted average of all available instrumental observations (magenta); and the ten largest volcanic eruptions in the 1400–2011 interval according to ice-core sulphate concentrations³⁰ (vertical red). **b**, As in **a**, but individual ensemble

minima, which are clustered in times following volcanic eruptions, as well as during the cool 1600s. As observed elsewhere²⁴, incidences of record-breaking temperature maxima have been increasing in frequency over the past few decades, whereas the rate of minima in the summer average temperature has fallen close to zero.

At a more regional level, the summer of 2010 is very likely ($P > 0.90$) to have been the warmest in seven of the eight grid boxes in the region affected by the Russian heat wave (50° N–60° N, 35° E–55° E (ref. 1); Fig. 2a). Averaging temperatures over this region reduces the variability between ensemble members, and we then find that 2010 is the warmest summer in western Russia in more than 99.9% of the ensemble members. In contrast, the summer of 2003 is probably ($P > 0.66$) the warmest for only two grid boxes in western Europe (Fig. 2a), in agreement with other findings^{2,9} that the 2010 Russian heat wave exhibited a broader spatial signature of extremes. We also find that 30 of the 47 continental grid boxes in western Greenland and the Canadian Arctic archipelago are more likely than not to have experienced their warmest conditions in 2010, in agreement with other more localized indications of temperature, runoff and surface mass balance in this region^{5–7}. For the average over this region, 2010 has the warmest conditions in 93% of the ensemble members.

The robustness of these results is indicated by the fact that using subsets of the proxy data gives similar structure with respect to the occurrence of extreme events. For instance, estimates based on only the ice-core and varve proxies, with model parameters derived using the complete data set (see Methods), give very similar results to inferences from all data sources, which are dominated by tree-ring densities

members are first smoothed with a nine-year Hanning window, along with a separate reconstruction¹⁰ (dashed red). **c**, As in **a**, but for linear trends calculated for overlapping 100-year intervals. To facilitate comparison between the recent rate of warming and earlier rates of cooling, the median and 90% pointwise uncertainty for the cooling centred at 1596 is also inverted in sign and extended across to the modern period (red lines and shading).

before 1850 and by instrumental observations thereafter (Fig. 2c). Both reconstructions show substantially elevated probabilities of maxima occurring in recent decades, very low probability of maxima in 1600–1850 and slightly elevated probabilities in the first two centuries of the reconstruction (Fig. 2b and c). Estimates using only tree-ring-density records result in a distribution of extreme years that is essentially unchanged, provided the comparison extends only to 1960, after which these proxies systematically underestimate temperature anomalies—as expected, given the tree-ring divergence phenomenon²⁵. More generally, we find that the various temperature indicators used in this study agree with each other given their respective uncertainties, excepting the tree-ring-density records after 1960 (see Methods and Supplementary Information).

The increased frequency of warm extremes in recent decades (Fig. 2b and c) may result from increased average temperature, increased variance or changes in higher-order moments of the temperature distribution²⁶. To explore the underlying cause of the recent increase in warm extremes, we use the Bayesian model and its data-derived parameters to simulate both temperature anomalies and observations of those anomalies having noise properties consistent with the instrumental data (see Methods). The distribution of simulated temperature anomalies over the past 20 years is biased substantially lower than estimates using the actual data (Fig. 3a). Shifting the mean of the simulated anomalies to match that observed over the past 20 years, which is 1.16 °C higher than the 600-year average, results in much closer agreement between the distributions of simulated and estimated temperatures (Fig. 3a), although the five highest individual

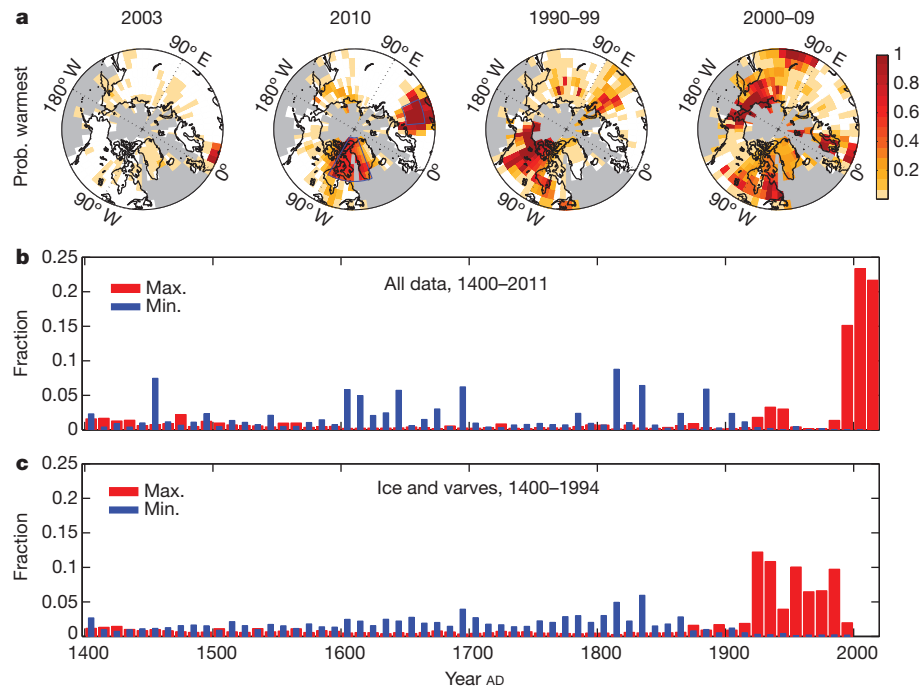


Figure 2 | Warm and cold extremes. **a**, The proportion of draws (see Methods) for which 2003 and 2010 were warmest, and for which the warmest year fell in the 1990s and 2000s. White shading indicates zero. **b**, The fraction of all locations for which years were warmest or coolest, averaged across all

ensemble members. Results are binned by decade, except for the last interval, which contains only 2010–2011. **c**, As in **b**, but for a reconstruction that uses only the ice-core and lake-varve series and covers the interval 1400–1994.

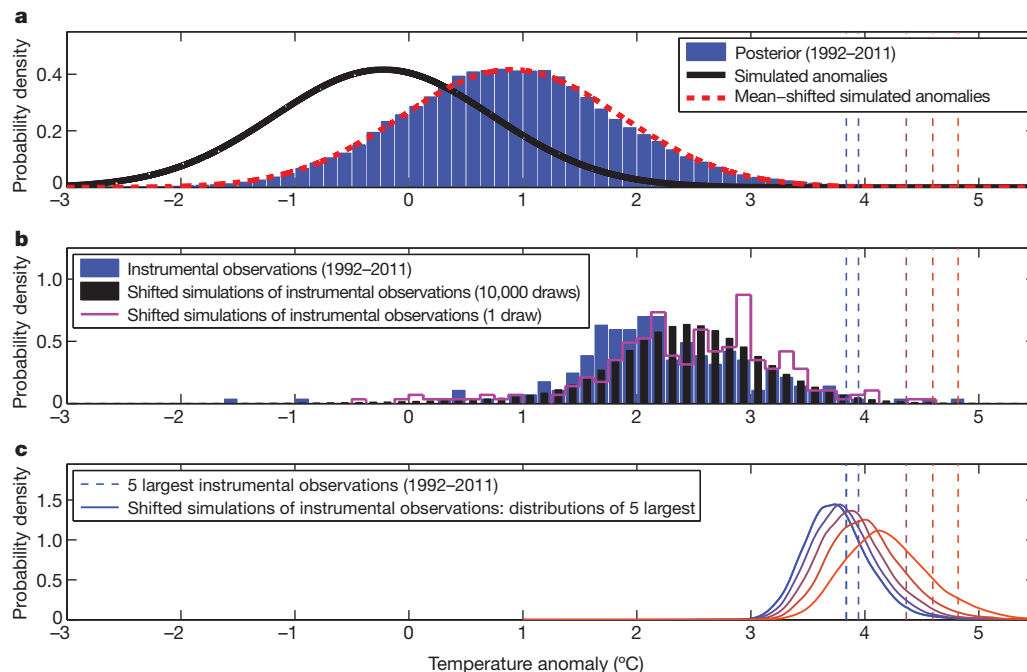


Figure 3 | Histograms of temperature anomalies and instrumental maxima for the period 1992–2011. **a**, Histogram of temperature anomalies across locations, ensemble members and years for the interval 1992–2011 (blue); the simulated distribution of temperature anomalies, using median parameter values fitted over 1400–2011 (black) and after shifting its mean to that of the 1992–2011 anomalies (dashed red). Vertical lines are repeated in each panel and correspond to the 3.8 °C anomaly near Moscow in 2010 and the four other even more extreme values in the 1992–2011 interval which, from lowest to

highest, are from Siberia in 2007, Svalbard in 2006 and two locations in northern Canada in 1998. **b**, Histogram of the maximum value of the instrumental observations at each location over the 1992–2011 interval (blue), and the distribution of the maximum values according to 10,000 realizations (black) and a single (magenta) realization based on the simulations with the mean shifted as in **a**. **c**, Distributions of the five largest simulated instrumental observations across space and time in the 1992–2011 interval, with the mean shifted as in **a**.

instrumental anomalies during the past 20 years remain far in the tail of the simulated distribution.

Analogous to the need for pairwise uncertainties when evaluating whether particular years featured the warmest conditions (Fig. 1), we must also take into account that the five highest observed temperature extremes are selected from a range of locations. Shifting the mean and selecting the largest simulated instrumental observations in both time (Fig. 3b) and space (Fig. 3c) results in distributions consistent with the observed extremes, each of which falls below the 95th percentile of the corresponding distribution from the simulations (Fig. 3c). To first order, therefore, the highest temperature observations over the past 20 years can be accounted for by a shift in the average temperature and by considering the distribution of maxima across both space and time.

We find a similar relationship between the mean and extremes when performing the same analysis for earlier 20-year intervals between 1852 and 1991 (see Supplementary Information), indicating that the observed extremes require no change to the distribution of temperature anomalies other than to the mean, at least on the temporal and spatial scales of this analysis. Applying the same analysis to 100-year intervals of each proxy data set gives similar results, although with more scatter than the results shown in Fig. 3. There are two exceptions where the distributions of positive extremes do not follow the mean: over the most recent 100-year interval of tree-ring-density observations, positive extremes are biased low on account of the divergence phenomenon, and there is a general tendency for the largest log-transformed varve observations to be biased high relative to the mean-shifted simulations, suggesting that the log transform is an insufficient scaling for these largest values.

During the past 20 years, 61% of all high northern latitude locations are more likely than not to have experienced warm temperatures that are unprecedented in the past 600 years, and 25% of northern locations are very likely ($P > 0.95$) to have experienced high temperatures that are without precedent. Assessment of these recent temperature extremes in the context of space–time variability and observational error allows for a simple, cohesive explanation in terms of an increase in mean temperature. Specifically, the Bayesian model of variability in conjunction with a mean warming reproduces the general distribution of temperature anomalies (Fig. 3a), the maximum instrumental observation at each location (Fig. 3b) and the highest instrumental temperature observations over space and time (Fig. 3c). These results are consistent with analyses of global instrumental records demonstrating that recent incidences of extreme temperature are a result of an upward shift in the mean of the temperature distribution²⁴ but that the variance of that distribution has not significantly changed²⁷. Further work is needed, however, to assess whether extremes also follow mean temperatures at smaller temporal and spatial scales, such as those associated with local weather.

METHODS SUMMARY

Observations are first re-expressed as anomalies with respect to the full time interval spanned by each data type²⁸ and the ice-core and log-transformed lake-varve series are additionally standardized to unit variance. The Bayesian model^{13,14} represents the temperature anomalies as a first-order autoregressive process in time with exponentially decaying spatial covariance. Each proxy type is separately modelled as linearly related to temperature anomalies with mean-zero normal noise, and instrumental observations are modelled as temperature anomalies with mean-zero normal noise.

The spatial mean time series is computed for each ensemble member by weighting the grid boxes by land area. Taking the 50th percentile at each year gives the best estimate of temperature, taking the 5th and 95th percentiles produces 90% pointwise uncertainty envelopes, and expanding the pointwise uncertainty envelopes to contain 90% of ensemble members in their entirety produces pairwise uncertainties^{15,17,29} (Fig. 1). Estimates of temporally smoothed quantities and centennial-scale trends are derived analogously after applying these operations to each ensemble member in turn (Fig. 1). Probability statements concerning extremes are derived by calculating the fraction of ensemble members that feature warmest or coldest conditions for a given year or interval (Fig. 2). To assess the robustness of results, we perform reconstructions using only subsets of the data

(Fig. 2c). To explore recent extremes, we simulate the expected variability of instrument-like observations in the absence of data (Fig. 3), using model parameters that result from the full analysis.

The analysis technique gives results at a level comparable to or better than other climate field reconstruction techniques²², even when the underlying assumptions are not fully met¹⁴. Validation metrics similarly suggest that the simple assumptions underlying the analysis are an adequate statistical description of the data, whereas experiments based on predicting withheld instrumental observations from the proxies indicate that the uncertainty intervals have accurate coverage properties.

Full Methods and any associated references are available in the online version of the paper.

Received 1 September 2012; accepted 29 January 2013.

- Dole, R. *et al.* Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.* **38**, L06702 (2011).
- Barriopedro, D., Fischer, E., Luterbacher, J., Trigo, R. & García-Herrera, R. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* **332**, 220–224 (2011).
- Rahmstorf, S. & Coumou, D. Increase of extreme events in a warming world. *Proc. Natl Acad. Sci. USA* **108**, 17905–17909 (2011).
- Otto, F. E. L., Massey, N., Oldenborgh, G. J. V., Jones, R. G. & Allen, M. R. Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophys. Res. Lett.* **39**, L04702 (2012).
- van As, D. *et al.* Large surface meltwater discharge from the Kangerlussuaq sector of the Greenland ice sheet during the record-warm year 2010 explained by detailed energy balance observations. *Cryosphere* **6**, 199–209 (2012).
- Tedesco, M. *et al.* The role of albedo and accumulation in the 2010 melting record in Greenland. *Environ. Res. Lett.* **6**, 014005 (2011).
- Richter-Menge, J. & Jeffries, M. The Arctic. *Bull. Am. Meteorol. Soc.* **92**, S143–S160 (2011).
- Overpeck, J. *et al.* Arctic environmental change of the last four centuries. *Science* **278**, 1251–1256 (1997).
- Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M. & Wanner, H. European seasonal and annual temperature variability, trends, and extremes since 1500. *Science* **303**, 1499–1503 (2004).
- Kaufman, D. S. *et al.* Recent warming reverses long-term Arctic cooling. *Science* **325**, 1236–1239 (2009); correction **327**, 644 (2010).
- National Research Council. *Surface Temperature Reconstructions for the Last 2000 Years* (National Academies Press, 2006).
- Tingley, M. *et al.* Piecing together the past: statistical insights into paleoclimatic reconstructions. *Quat. Sci. Rev.* **35**, 1–22 (2012).
- Tingley, M. & Huybers, P. A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 1: development and applications to paleoclimate reconstruction problems. *J. Clim.* **23**, 2759–2781 (2010).
- Tingley, M. & Huybers, P. A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 2: comparison with the regularized expectation-maximization algorithm. *J. Clim.* **23**, 2782–2800 (2010).
- Craigmile, P., Guttorp, P. & Percival, D. Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics* **15**, 313–335 (2004).
- Li, B., Nychka, D. & Ammann, C. The ‘hockey stick’ and the 1990s: a statistical perspective on reconstructing hemispheric temperatures. *Tellus A* **59**, 591–598 (2007).
- McShane, B. & Wyner, A. A statistical analysis of multiple temperature proxies: are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Stat.* **5**, 5–44 (2011).
- Christiansen, B. & Ljungqvist, F. Reconstruction of the extra-tropical NH mean temperature over the last millennium with a method that preserves low-frequency variability. *J. Clim.* **24**, 6013–6034 (2011).
- Christiansen, B. & Ljungqvist, F. The extra-tropical northern hemisphere temperature in the last two millennia: reconstructions of low-frequency variability. *Clim. Past* **8**, 765–786 (2012).
- Gelman, A., Carlin, J., Stern, H. & Rubin, D. *Bayesian Data Analysis* 2nd edn (Chapman and Hall/CRC, 2003).
- Berliner, L., Wikle, C. & Cressie, N. Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *J. Clim.* **13**, 3953–3968 (2000).
- Werner, J., Luterbacher, J. & Smerdon, J. A pseudoproxy evaluation of Bayesian hierarchical modeling and canonical correlation analysis for climate field reconstructions over Europe. *J. Clim.* **26**, 851–867 (2013).
- Brohan, P., Kennedy, J., Harris, I., Tett, S. & Jones, P. Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. *J. Geophys. Res.* **2**, 99–113 (2006).
- Hansen, J., Sato, M. & Ruedy, R. Perception of climate change. *Proc. Natl Acad. Sci. USA* **109**, E2415–E2423 (2012).
- Briffa, K. *et al.* Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. *Nature* **391**, 678–682 (1998).
- IPCC. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. *et al.*) 1–19 (A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change, Cambridge Univ. Press, 2012).
- Rhines, A. & Huybers, P. Frequent summer temperature extremes reflect changes in the mean, not the variance. *Proc. Natl Acad. Sci. USA* **110**, E546 (2013).

28. Tingley, M. A Bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record. *J. Clim.* **25**, 777–791 (2012).
29. Krivobokova, T., Kneib, T. & Claeskens, G. Simultaneous confidence bands for penalized spline estimators. *J. Am. Stat. Assoc.* **105**, 852–863 (2010).
30. Gao, C., Robock, A. & Ammann, C. Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. *J. Geophys. Res. D* **113**, D23111 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements Data analysis was performed on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University. Funding

for this work was provided in part by NSF grant ATM-0902374. We thank E. Butler, P. Craigmile, N. Cressie, M. Haran, B. Li, E. Mannshardt, K. McKinnon, D. Nychka, B. Rajaratnam, A. Rhines, D. Schrag and A. Stine for comments and discussions.

Author Contributions M.P.T. performed the analysis. Both authors contributed to the design of the analysis, the interpretation of results, and preparation of the manuscript.

Author Information Data and code are available from NOAA Paleoclimatology at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/contributions_by_author/tingley2013/. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.P.T. (tingley@fas.harvard.edu).

METHODS

Data. Instrumental observations are derived from the gridded $5^\circ \times 5^\circ$ CRUTEM3v data set compiled by the University of East Anglia's Climate Research Unit (ref. 23; data available at <http://www.cru.uea.ac.uk/cru/data/temperature/>). We use all time series having at least ten years of complete monthly April–September data, that are poleward of 45° N, and that contain a non-zero fraction of land according to a $0.25^\circ \times 0.25^\circ$ land–sea mask (ref. 31; data available at <http://ldas.gsfc.nasa.gov/gldas/GLDASVegetation.php>). To avoid introducing spurious structure in the time series of spatial standard deviations, which could result from the short 1961–90 reference period used in standardizing the CRUTEM3v data set, we apply a Bayesian ANOVA technique (ref. 28; code available at <http://www.ncdc.noaa.gov/paleo/softlib/softlib.html>) to estimate and remove means with respect to the entire 1850–2011 interval spanned by the instrumental data set.

The maximum latewood-density data set^{32,33} is on the same spatial grid as the instrumental data set, and we use only the 96 grid boxes with centres poleward of 45° N. Data files and descriptions are available at <http://www.cru.uea.ac.uk/~timo/datapages/mxdrtrw.htm>. As with the instrumental data, we apply the Bayesian ANOVA technique²⁸ to estimate and remove means with respect to the entire interval spanned by the data set, in this case 1400–1994.

All varve thickness records publicly available from the NOAA Paleolimnology Data Archive (http://www.ncdc.noaa.gov/paleo/paleolim/paleolim_data.html) as of January 2012 are incorporated, provided they meet the following criteria: extend back at least 200 years, are at annual resolution, are reported in length units, and the original publication or other references indicate or argue for a positive association with summer temperature. As is common³⁴, varve thicknesses are logarithmically transformed before analysis, giving distributions that are more nearly normally distributed and in agreement with the assumptions characterizing our analysis (see subsequent section). Records that are complete between 1400 and 1969 are standardized to unit variance and zero mean using the sample statistics computed over that interval. To obtain a more homogeneous normalization for records incomplete between 1400 and 1969, we scale the mean and standard deviation of each incomplete record to equal those statistics of the corresponding data points in the scaled, complete records. Finally, the Bayesian ANOVA technique²⁸ is used to remove means with respect to the entire 1400–2005 time span of the varve data set.

The ice-core data set consists of 14 of the 15 annually resolved $\delta^{18}\text{O}$ series used in a recent sea ice reconstruction³⁵. We exclude the Mount Logan series, as the original reference³⁶ indicates it is a proxy for precipitation source region, not temperature. To increase the spatial coverage, we additionally use two $\delta^{18}\text{O}$ series from Svalbard (refs 37, 38; data available at <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/polar/svalbard/svalbard2005d18o.txt>) and one each from Baffin Island (refs 39, 40; data file `fisher_1998_baffin.ppd` available at <http://www.ncdc.noaa.gov/paleo/pubs/pcn/pcn-proxy.html>) and Devon Island (ref. 41; data available at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/polar/devon/d7273del_1yr.txt), all with annual observations. The data set spans the 1400–1998 interval; the standardization procedure is the same as for the log-transformed varve series except that it uses the 1400–1974 interval for initial standardization.

Section 1 of Supplementary Information and Supplementary Tables 1 and 2 provide additional details regarding the data sources.

Inference. Bayesian hierarchical models provide a flexible framework for combining models and observations²⁰. They are generally characterized by a process level that represents the structure of the system and a data level that represents the relationships between each data type and the process. Both the process and data levels contain parameters that are uncertain and whose distributions are inferred as part of the analysis. Similarly, each observation is considered uncertain; in the current analysis this includes both the proxy and the instrumental observations. The process targeted for inference is, therefore, never directly observed but must be inferred using uncertain observations and a model with parameters whose distributions must likewise be estimated. Although Bayesian hierarchical models have been proposed and used to infer past climate^{12–14,42,43}, the analysis presented here is, to our knowledge, the first application of such a model to infer surface temperatures as a function of space and time from a multiproxy data set.

In this analysis, we use the Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time (BARCAST)^{13,14}. In our application, the process level describes the evolution of the average April–September temperature anomaly field, T , as,

$$T_t - \mu I = \alpha(T_{t-1} - \mu I) + \epsilon_t \quad (1)$$

where μ is the mean of the temperature anomaly field, α is the coefficient of a first-order autoregressive process, the subscript t indexes the year, and I is a column vector of ones. The temperature field is represented on a $5^\circ \times 5^\circ$ grid, and we consider only continental locations that are north of 45° N. The innovation vector

for each year, ϵ_t , is assumed to be an independent draw from a mean-zero multivariate normal distribution with an exponentially decaying spatial correlation: $\epsilon_t \sim N(0, \Sigma)$, where $\Sigma_{ij} = \sigma^2 \exp(-\phi |s_i - s_j|)$, and $|s_i - s_j|$ is the distance between the i th and j th elements of the field vector, T_t .

The data level describes the relationships between the true temperature anomalies and the instrumental and proxy observations of these anomalies. Instrumental observations, $W_{0,t}$, are modelled as noisy versions of the true anomalies at the corresponding locations,

$$W_{0,t} = H_{0,t} T_t + e_{0,t} \quad (2)$$

The noise terms are assumed to be independent draws from a multivariate normal distribution, $e_{0,t} \sim N(0, \tau_0^2 I)$, where I is the identity matrix, and $H_{0,t}$ is a selection matrix of zeros and ones that picks out elements of T_t for which there are instrumental observations.

The types of proxy observations included are tree-ring density chronologies, ice-core $\delta^{18}\text{O}$ series, and log-transformed lake-varve thickness series. Each type, k , is assumed to have a linear relationship with the local value of the true temperature,

$$W_{k,t} = \beta_{k,1} H_{k,T} T_t + \beta_{k,0} I + e_{k,t} \quad (3)$$

where $\beta_{k,1}$ and $\beta_{k,0}$ are respectively the slope and intercept terms and $H_{k,t}$ is, as above, a selection matrix. The noise terms are once more assumed to be independent normal draws, $e_{k,t} \sim N(0, \tau_k^2 I)$. The regression parameters vary between proxy types, but are constant for all proxies of a given type.

Prior distributions are placed on each of the parameters included in the model: $\alpha, \mu, \phi, \sigma^2, \tau_k^2, \beta_{k=1 \dots 3,1}, \beta_{k=1 \dots 3,0}$, as well as for T in the year before observations become available. Priors are selected to be proper, weakly informative, and—where possible—conjugate²⁰. Bayes' rule is used to calculate the posterior distribution of the parameters and field given the observations and priors. A Gibbs sampler with a single Metropolis step (for ϕ) is then used to draw from the posterior distribution²⁰. Further details of the inference are available in an earlier publication¹³, code is available at <http://www.ncdc.noaa.gov/paleo/softlib/softlib.html>, and convergence of the Gibbs sampler is discussed in Section 3.5 of the Supplementary Information.

The result of the analysis is an ensemble of draws of both the parameters and the temperature anomalies, each of which is equally likely given the data, priors and modelling assumptions. Furthermore, each ensemble member will have variability similar to the actual temperature anomalies¹⁴, inasmuch as the model and data are correctly represented. The median across the ensemble (Fig 1a, b; Supplementary Fig. 5), however, has attenuated variability¹⁴, especially in data-poor parts of the reconstruction. This attenuation provides a more accurate estimate of the past temperature, though not the variability of that temperature, and is generally used as our best estimate.

For the purposes of comparing the climatic information content of the different proxies (for example, Fig. 2c), it is possible to run BARCAST in a reduced mode, using a subset of the data to update the temperature field while sampling all model parameters from the posterior distributions resulting from the analysis with the full data set. That is, at each iteration of the Gibbs sampler, a vector of parameters is drawn from the posterior distribution derived using all data, and then the draw of the temperature field is updated using these parameters and a given subset of the data. It is likewise possible to simulate the natural variability of the temperature field (for example, Fig. 3) by not applying any of the data constraints. In this case, the process level model has parameters constrained by the data but the specific evolution of the temperature field is unconstrained. Comparison of the variability between the constrained and unconstrained simulations indicates the extent to which the data controls the solution (see Supplementary Information Sections 2.5 and 6).

Assumptions and implications. The stationary, isotropic and exponentially decaying spatial covariance model specified for the temperature anomalies is a simplifying assumption that does not account for directionality and long-range covariance relationships in the climate system. Indeed, many palaeoclimate reconstruction techniques, generally based on eigendecompositions of sample covariance or correlation matrices, are explicitly designed to exploit such covariance patterns^{9,44–46}. This class of methods assumes that the characteristic spatial structures identified in the calibration interval are constant in time, but have varying amplitudes; results can be strongly dependent on the particulars of how these modes of variability are determined and used⁴⁷. BARCAST, in contrast, relies on the simpler assumption of a temporally constant decorrelation length scale and represents each observation as indicative of local temperature. This local representation (equation (3)) is similar to the proxy representation in the LOC method^{18,19,48}, although BARCAST additionally models errors in the instrumental observations and arrives at an estimate of the spatial field of temperature anomalies^{49,50}.

Temporal stationarity is assumed through the first-order autoregressive description of interannual temperature variability (equation (1)). This parameterization lacks any representation of long-term temperature trends, or associations between temperatures and climate forcings^{43,51}. In some sense, however, this is advantageous as the process level is agnostic regarding any changes, and inferences concerning the unprecedented nature of recent extremes are consequently derived exclusively from the data. Furthermore, this approach permits exploration of exactly where the assumption of strict temporal stationarity fails (see Fig. 3).

The assumption that the space–time covariance function is separable—that is, that it can be factored into the product of purely spatial and purely temporal elements—is also unlikely to hold in detail¹². Predictive performance, however, is often not affected by incorrectly assuming a separable covariance form⁵². With regard to both stationarity and separability, BARCAST has been shown to perform at a level comparable to or better than other climate field reconstruction techniques²², even when the underlying assumptions are not strictly met by the data¹⁴. The validation metrics reported in Supplementary Information Section 5 also suggest that BARCAST provides an adequate statistical description of the data for the present purposes.

Analysis of results. The analysis presented here is based on 4,000 posterior draws taken from four parallel Gibbs samplers. Spatial mean values are computed by weighting the grid boxes by land area according to a 0.25° land–sea mask³¹. Taking the 50th percentile of the 4,000-member ensemble at each year of the spatial mean time series results in a best estimate of the time series, while taking the 5th and 95th percentiles produces 90% pointwise credible intervals²⁰, which are used to indicate the uncertainty in the reconstruction at each year.

Pathwise 90% credible intervals are calculated by inflating the pointwise intervals such that the envelope contains 90% of the posterior time series in their entirety^{15,17,29}. The two uncertainty intervals have different interpretations, with the pointwise intervals covering the true temperature anomaly for 90% of the years, whereas the time series of true temperature anomalies lies entirely within the pathwise envelope with 90% probability. The statement that certain recent years are warmer than all previous years with $P > 0.95$ follows from noting that the 90% pathwise uncertainties for these years lie entirely above those for all years in the 1400–2004 interval. Note that the statement holds at $P > 0.95$ because the test is one sided, as we have prior reason to believe that recent years are warmer than usual. Uncertainties for temporally smoothed time series are estimated in the same manner, after first smoothing each ensemble member using a nine-point Hanning window. Point estimates and uncertainties in the centennial-scale trends are derived by calculating the trend for each ensemble member at each position of a sliding 100-year window and then calculating both pointwise and pathwise uncertainties.

Probability estimates corresponding to specific statements, such as the probability that western Russian temperatures achieved a maximum in 2010, are obtained by calculating the proportion of ensemble members for which the statement is true. Performing such an analysis at each location leads to the maps shown in Fig. 2a. Calculating the fraction of locations for which each year is warmest or coldest, averaging results across the ensemble members and binning the years by decades results in the histograms shown Fig. 2b. Figure 2c is derived in the same manner, but from a 4,000-member ensemble obtained by running BARCAST in reduced mode using only the ice-core and log-transformed lake-varve series.

To assess recent extremes, we simulate temperature anomalies and instrument-like observations over the past 20 years using median parameter values, and in the latter case additionally record the maxima at each location as well as the five largest values in both space and time. Distributions of the simulated instrumental quantities are built up by repeating the procedure 10,000 times. When shifting the mean of these simulations to match that inferred over the past 20 years (Fig. 3), we use the simple average across the locations, ensemble members, and years within the 1992–2011 interval. Note that this mean differs slightly from the mean of the spatial average time series because of the spatial weighting inherent in the latter. Figure 3b shows the single instrumental simulation that is closest to the actual instrumental observations according to the variance of the site-wise maxima. Additional results and model diagnostics are available in Supplementary Information Sections 2 and 3 and a more detailed discussion on inferring extremes in the presence of uncertainty in Supplementary Information Section 6.

Robustness. To examine the robustness of our results to specific data types, we run BARCAST in five different reduced mode formulations using the following subsets of data types: tree-ring densities alone, ice-core series alone, lake-varve series alone, ice-core and lake-varve series together, and instrumental series alone. Results of the proxy-only analyses are then compared with each other, to the main analysis, to the instrumental-only analysis, and to the withheld instrumental observations. The time series of spatial-average temperature and centennial-scale slopes, as well as the distribution of extremes for the spatial mean, for centennial

trends in the spatial mean, and as a function of space, are all compared with each other. For the spatial average, we also consider the correlations and root-mean-square error between each proxy-only analysis and the instrumental-only analysis over three different time intervals: 1850–1959, 1850–1994 and 1960–94. In general, we find that the proxy-only analyses provide consistent inferences with one another and with the instrumental-only predictions, when accounting for the uncertainties in each analysis (see Supplementary Information Sections 5.1–5.4).

An important exception is that the predictions from the tree-ring densities alone do not track the warming seen in all other data sets in the latter half of the twentieth century, a finding consistent with the so-called divergence problem^{33,53}. To assess the robustness of our results to this divergence, we re-ran the full analysis excluding the post-1960 tree-ring-density observations and found no qualitative change in results (see Supplementary Information Section 4). The primary reason for consistency between analyses which include and exclude the post-1960 tree-ring-density observations is that instrumental data are of sufficient quality and number post-1960 so as to dominate the solution irrespective of the tree-ring data. It is also the case that the parameterization of the tree-ring-density relationship with temperature is primarily constrained by its relationship with data before 1960.

To assess the variability in the ensemble of posterior draws, we simulate the withheld instrumental observations in each proxy-only analysis and examine rank verification histograms⁵⁴ for the withheld instrumental observations. To assess the coverage rates of the credible intervals, we calculate the percentage of the withheld instrumental observations that fall within the nominal 90% intervals. Results show that the ensembles of predictions of the withheld instrumental values generally have about the correct variability, and that the actual coverage rates are generally within 10% of the nominal rate. An exception is, again, for the tree-ring-density analysis during the post-1960 interval, where the coverage rate is about 15% too low and the shape of the rank verification histogram is indicative of a low bias. See Supplementary Information Section 5.5 for further details.

- Rodell, M. *et al.* The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **85**, 381–394 (2004).
- Briffa, K. *et al.* Tree-ring width and density data around the northern hemisphere: part 1, local and regional climate signals. *Holocene* **12**, 737–757 (2002).
- Briffa, K. *et al.* Tree-ring width and density data around the northern hemisphere: part 2, spatio-temporal variability and associated climate patterns. *Holocene* **12**, 759–789 (2002).
- Loso, M. Summer temperatures during the medieval warm period and little ice age inferred from varved proglacial lake sediments in southern Alaska. *J. Paleolimnol.* **41**, 117–128 (2009).
- Kinnard, C. *et al.* Reconstructed changes in Arctic sea ice over the past 1,450 years. *Nature* **479**, 509–512 (2011).
- Fisher, D. *et al.* The Mt Logan Holocene–late Wisconsinan isotope record: tropical Pacific–Yukon connections. *Holocene* **18**, 667–677 (2008).
- Isaksson, E. *et al.* Climate oscillations as recorded in Svalbard ice core $\delta^{18}\text{O}$ records between 1200–1997 AD. *Geogr. Ann.* **87**, 203–214 (2005).
- Isaksson, E. *et al.* Svalbard Ice Cores 600 Year Annual $\delta^{18}\text{O}$ Data (IGBP PAGES/World Data Center for Paleoclimatology, Data Contribution Series no. 2011-068. NOAA/NCDC Paleoclimatology Program, Boulder 2011); available online at <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/polar/svalbard/svalbard2005d18o.txt>.
- Fisher, D. *et al.* Penny ice cap cores, Baffin Island, Canada, and the Wisconsinan Foxe Dome connection: two states of Hudson Bay ice cover. *Science* **279**, 692–695 (1998).
- Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105**, 13252–13257 (2008).
- Fisher, D. Comparison of 105 years of oxygen isotope and insoluble impurity profiles from the Devon Island and Camp Century ice cores. *Quat. Res.* **11**, 299–305 (1979).
- Haslett, J. *et al.* Bayesian palaeoclimate reconstruction. *J. R. Stat. Soc. A* **169**, 395–438 (2006).
- Li, B., Nychka, D. & Ammann, C. The value of multi-proxy reconstruction of past climate. *J. Am. Stat. Assoc.* **105**, 883–895 (2010).
- Mann, M., Bradley, R. & Hughes, M. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**, 779–787 (1998).
- Schneider, T. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J. Clim.* **14**, 853–871 (2001).
- Smerdon, J., Kaplan, A., Chang, D. & Evans, M. A pseudoproxy evaluation of the CCA and RegEM methods for reconstructing climate fields of the last millennium. *J. Clim.* **23**, 4856–4880 (2010).
- Christiansen, B., Schmith, T. & Thejll, P. A surrogate ensemble study of sea level reconstructions. *J. Clim.* **23**, 4306–4326 (2010).
- Christiansen, B. Reconstructing the NH mean temperature: can underestimation of trends and variability be avoided? *J. Clim.* **24**, 674–692 (2011).

49. Tingley, M. & Li, B. Comments on 'Reconstructing the NH mean temperature: can underestimation of trends and variability be avoided?'. *J. Clim.* **25**, 3441–3446 (2012).
50. Christiansen, B. Reply to "Comments on 'Reconstructing the NH mean temperature: can underestimation of trends and variability be avoided?'". *J. Clim.* **25**, 3447–3452 (2012).
51. Lee, T., Zwiers, F. & Tsao, M. Evaluation of proxy-based millennial reconstruction methods. *Clim. Dyn.* **31**, 263–281 (2008).
52. Genton, M. Separable approximations of space–time covariance matrices. *Environmetrics* **18**, 681–695 (2007).
53. D'Arrigo, R., Wilson, R., Liepert, B. & Cherubini, P. On the 'Divergence problem' in northern forests: a review of the tree-ring evidence and possible causes. *Glob. Planet. Change* **60**, 289–305 (2008).
54. Hamill, T. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weath. Rev.* **129**, 550–560 (2001).

Climatic control of bedrock river incision

Ken L. Ferrier¹†, Kimberly L. Huppert¹ & J. Taylor Perron¹

Bedrock river incision drives the development of much of Earth's surface topography^{1–3}, and thereby shapes the structure of mountain belts⁴ and modulates Earth's habitability through its effects on soil erosion⁵, nutrient fluxes⁶ and global climate⁷. Although it has long been expected that river incision rates should depend strongly on precipitation rates, quantifying the effects of precipitation rates on bedrock river incision rates has proved difficult, partly because river incision rates are difficult to measure and partly because non-climatic factors can obscure climatic effects at sites where river incision rates have been measured^{8,9}. Here we present measurements of river incision rates across one of Earth's steepest rainfall gradients, which show that precipitation rates do indeed influence long-term bedrock river incision rates. We apply a widely used empirical law for bedrock river incision^{3,9–11} to a series of rivers on the Hawaiian island of Kaua'i, where mean annual precipitation ranges from 0.5 metres to 9.5 metres (ref. 12)—over 70 per cent of the global range¹³—and river incision rates averaged over millions of years can be inferred from the depth of river canyons and the age of the volcanic bedrock. Both a time-averaged analysis and numerical modelling of transient river incision reveal that the long-term efficiency of bedrock river incision across Kaua'i is positively correlated with upstream-averaged mean annual precipitation rates. We provide theoretical context for this result by demonstrating that our measurements are consistent with a linear dependence of river incision rates on stream power, the rate of energy expenditure by the flow on the riverbed. These observations provide rare empirical evidence for the long-proposed coupling between climate and river incision, suggesting that previously proposed feedbacks among topography, climate and tectonics may occur.

Anyone who has stood next to a river during a rainstorm can readily appreciate that precipitation plays an important part in regulating erosion rates. Precipitation facilitates river incision and sediment transport by increasing water discharge; it triggers landslides by elevating soil pore pressures; it accelerates the production of erodible soil by promoting chemical and physical weathering; and it fosters the growth of biota, which transport sediment and break down rock. The influence of precipitation rates on erosion rates seems so likely that spatially variable precipitation has been proposed to regulate the deformation, width and asymmetry of entire mountain ranges^{4,14}. It is therefore surprising that, despite theoretical advances in understanding the mechanisms by which water fluxes promote erosion, there remains no empirical consensus on how precipitation rates affect erosion rates in nature^{15–19}. Some studies have found no relationship between erosion rates and precipitation rates²⁰, whereas others have proposed functional relationships between erosion rates and precipitation rates that are highly inconsistent with one another²¹.

These inconsistencies have several possible roots. First, non-climatic factors such as tectonic uplift rates and rock type vary among sites where erosion rates have been measured, and these confounding factors can obscure climatic controls on erosion rates. Second, the effects of precipitation rates on erosion rates could be weak, and hence difficult to observe at sites that do not span wide ranges in precipitation rates. Third, negative feedbacks between precipitation and erosion

may emerge at higher precipitation rates if, for instance, sediment protects the riverbed or denser vegetation inhibits hillslope erosion. Fourth, many studies have focused on basin-averaged erosion rates, which reflect the sum of many erosion processes, each with its own characteristic response to precipitation rates. Although such studies provide important insights into the integrated effects of precipitation rates on basin-averaged sediment fluxes²², they suggest that a clearer picture may emerge by studying a single erosion process across a study area in which little varies except the precipitation rate. That is our approach here.

Here we focus on the effects of precipitation rates on bedrock river incision, a process that drives the topographic evolution of dominantly erosional landscapes such as mountain ranges¹¹. We explore the influence of precipitation rates on river incision rates by applying a widely used river incision law to a series of bedrock river channels across a wide range of precipitation rates on the Hawaiian island of Kaua'i. The stream power law for bedrock river incision³ is based on the supposition that the river incision rate E (with units of $L T^{-1}$, where L is length and T is time) scales with the rate at which the flow expends energy on the riverbed per unit bed area, or the unit stream power, ω :

$$E = k_b \omega^a \quad (1)$$

where k_b is a dimensional coefficient and a is a dimensionless exponent. For steady, uniform flow, $\omega = \rho g Q S / W$, where ρ is the fluid density, g is the gravitational acceleration, Q is the stream discharge, S is the channel gradient and W is the channel width³.

To facilitate application to topographic data, the stream power law is often recast by assuming that W scales with Q (refs 23, 24) and that Q scales with upstream drainage area²⁵, A , which yields^{1,3,8–11}:

$$E = K A^m S^n \quad (2)$$

where K , m and n are constants. This simplification is also motivated in part by the empirical observations that channel erosion rates scale as a power function of slope and drainage area¹⁰ and that channel slope and drainage area are negatively correlated in many fluvial networks³. An expression with the same form as equation (2) results from an assumption that incision rate is proportional to the shear stress exerted by the flow on the riverbed^{10,11}. Precipitation rate does not appear in this simplified and widely applied form of the stream power law, but its effect is implicitly contained in the coefficient K . We therefore expect that K will vary with precipitation rate, but this prediction has not been systematically tested against field observations, partly because of the difficulty of measuring K and partly because of the difficulty of isolating the effects of precipitation from the numerous other factors that are thought to influence K , such as rock strength. We investigate the climatic control of bedrock river incision by measuring how K varies across one of Earth's steepest precipitation gradients.

We focus our study on bedrock rivers on Kaua'i, a roughly circular volcanic island with an approximate diameter of 50 km and a maximum elevation of 1,593 m (Fig. 1). Aside from carbonate beach sands along the coast, the surface rocks of Kaua'i are composed of tholeiitic

¹Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. †Present address: Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, 02138, USA.

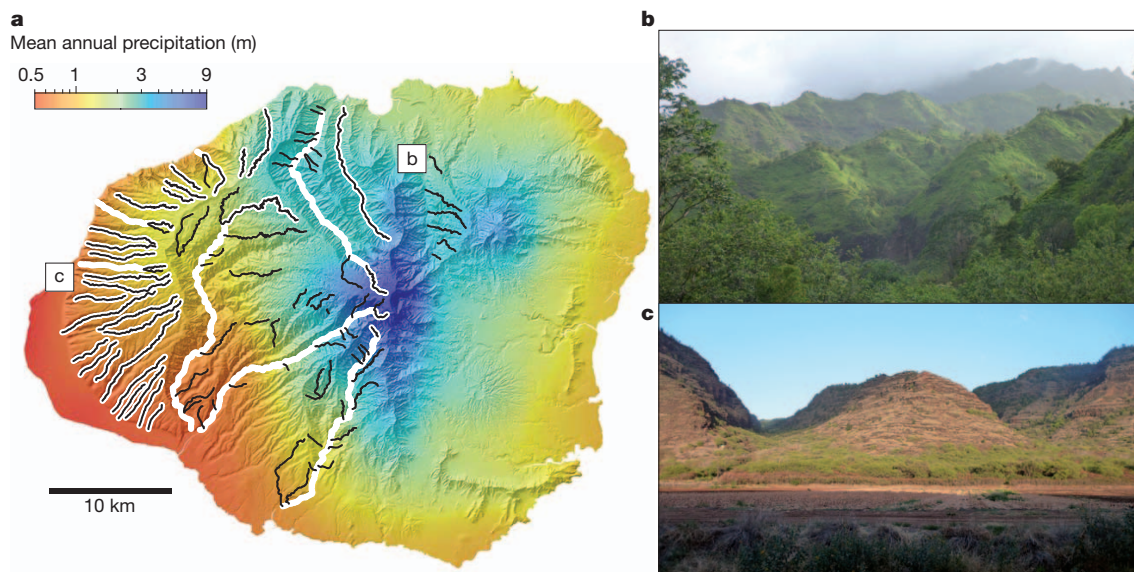


Figure 1 | Rainfall across Kaua'i. Colours show modern mean annual precipitation¹², resampled to 10-m resolution, overlain on a shaded relief map. White lines indicate channels examined in the time-averaged analysis, black

lines indicate channels examined in the transient analysis, and black lines on top of white lines indicate channels examined in both analyses. Photographs on the right show sites in the wetter and drier parts of Kaua'i.

or alkalic basalt with ages ranging from 5.1 to 0.15 million years (Myr) (refs 26–28) and consequently display relatively little lithologic variability across the island. Kaua'i also has one of Earth's largest gradients in mean annual precipitation rates. Between 1949 and 2004, precipitation rates averaged 9.5 m yr^{-1} atop Mt Wai'ale'ale near the centre of the island, whereas precipitation rates between 1949 and 2000 averaged only 0.5 m yr^{-1} on the southwestern part of the island^{29,30} (Fig. 1). That is, over a distance of just 25 km, precipitation rates on Kaua'i span >70% of the range in precipitation rates of the entire planet¹³. This extraordinarily wide range of precipitation rates and the minimal variations in lithology make Kaua'i an exceptional natural laboratory for investigating the effects of precipitation rates on bedrock river incision rates.

We used two methods based on equation (2) to investigate the effects of precipitation rates on bedrock river incision rates on Kaua'i. First, we calculated K for the time-averaged incision of 32 of Kaua'i's largest river valleys (Fig. 1; Methods); we refer to this as the time-averaged analysis. In the second approach, we modelled the transient evolution of 93 river channel segments on Kaua'i (Fig. 1) from their initial longitudinal profiles to their modern profiles to obtain best-fit values of K (Fig. 2; Methods); we refer to this as the transient analysis.

In the time-averaged analysis, we obtained estimates of E , A and S at 13,701 grid points in the 32 channels (Methods) and used these data to calculate best-fit values for the exponents m and n in equation (2) in a multiple linear regression: $\log_{10}(E) = \log_{10}(K) + m\log_{10}(A) + n\log_{10}(S)$. This yielded best-fit means and 95% confidence intervals of $m = 0.59 \pm 0.01$ and $n = 0.33 \pm 0.02$. We used these global best-fit values for m and n to calculate the time-averaged rate coefficient of bedrock river incision, $K = E/(A^m S^n)$, at each grid point.

In the transient analysis, we identified a best-fit value of K for each of the 93 river segments by minimizing the misfit between modelled and modern profiles (Fig. 2; Methods). The exponent n was assumed to be 1 for all profiles, consistent with an incision rate linearly proportional to stream power, and the exponent m was set to 0.59, the value found in the time-averaged analysis. A sensitivity analysis confirmed that the fits were relatively insensitive to the choice of n (Supplementary Information).

We then sought to determine how the efficiency of bedrock river incision depends on precipitation rate. Figure 3 plots the rate coefficient K , normalized by the mean K obtained in each analysis, against mean annual precipitation rate. The primary observation is that both the time-averaged analysis (Fig. 3a) and the transient analysis (Fig. 3b)

demonstrate a positive correlation between the efficiency of bedrock river incision and the upstream-averaged mean annual precipitation rate P . Moreover, the time-averaged and transient analyses yield similar best-fit power law relationships: erosional efficiency scales with $P^{0.59 \pm 0.01}$ in Fig. 3a and $P^{0.53 \pm 0.10}$ in Fig. 3b (mean \pm standard error, s.e.). Thus, even after accounting for the effects of channel slope and drainage area, river incision rates on Kaua'i exhibit a residual dependence on upstream precipitation rates. Although the modern precipitation rates we used to calculate P may differ from palaeoprecipitation rates, rainfall in the past was probably also controlled by a combination of trade winds and island topography³¹. Moreover, there are no obvious factors (such as rock type or tectonics) that might affect river incision rates and that also fortuitously covary with modern precipitation rates on Kaua'i (Supplementary Information). Thus, we suggest that the simplest explanation for the correlations in Fig. 3 is that mean annual precipitation rates strongly influence bedrock river incision rates.

The systematic increase in the coefficient K with mean annual precipitation rates on Kaua'i indicates that channel slope S and drainage area A alone are not sufficient predictors of river incision rates. Thus, when applied uniformly with a constant K , equation (2) does a poor job of predicting river incision rates in landscapes with significant rainfall gradients. A better approach in such landscapes is to calculate stream power directly from discharge using equation (1), which can be

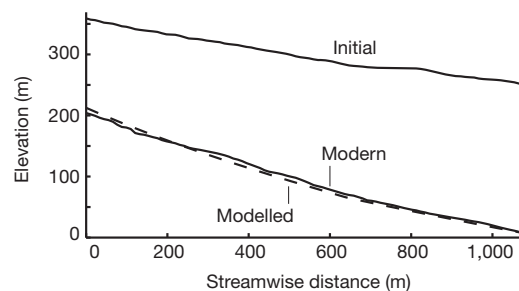


Figure 2 | Transient model of river incision. The Kulanaililia River's modern longitudinal profile (lower solid line), its estimated initial profile (upper solid line) and a best-fit modelled profile (dashed line) calculated by simulating the transient evolution of the channel from its initial profile to its modern profile (see Methods).

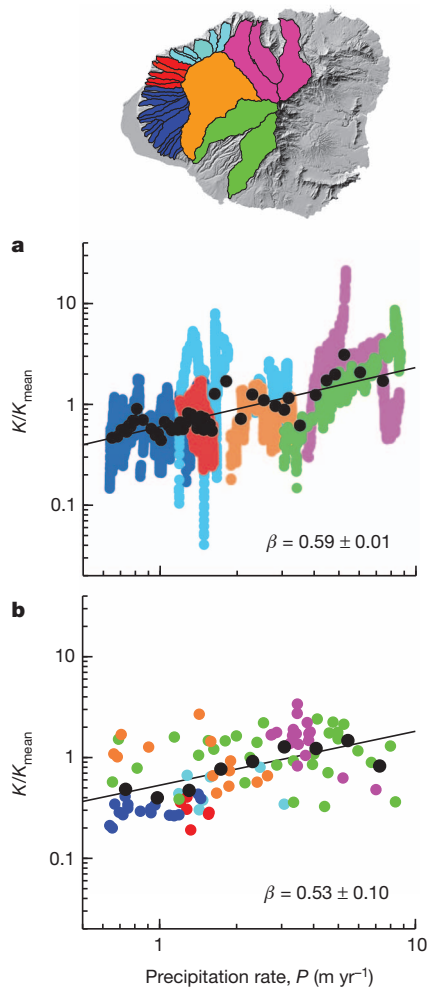


Figure 3 | Influence of precipitation rate on the efficiency of river incision.

a, Time-averaged analysis. Normalized river incision efficiency K/K_{mean} versus upstream-averaged mean annual precipitation rate P at each point. Data point colour indicates geographic region (see inset map). Black points are logarithmically binned averages. Here $K_{\text{mean}} = 1.53 \times 10^{-8} \text{ m}^{-0.18} \text{ yr}^{-1}$ and $R^2 = 0.33$ on the regression. **b**, Transient analysis. Normalized river incision efficiency versus basin-averaged P for each channel segment. Each data point represents the best-fit value of K , determined by modelling the transient evolution of a single channel from its initial profile to its present profile. Here $K_{\text{mean}} = 1.45 \times 10^{-7} \text{ m}^{-0.18} \text{ yr}^{-1}$ and $R^2 = 0.25$ on the regression. Regression slopes β (mean \pm s.e.) in both analyses suggest that bedrock river incision is more efficient at higher mean annual precipitation. Precipitation ranges for a given geographic region do not always overlap in **a** and **b** because the two analyses used different parts of the drainage network (Fig. 1). Differences in K_{mean} between **a** and **b** result in part from different n values in the two analyses.

straightforwardly related to precipitation rate, rather than using the drainage-area-based simplification in equation (2). In many landscapes, it is difficult or impossible to calculate stream power, owing to a lack of field measurements. On Kaua'i, however, measurements of discharge and channel width at United States Geological Survey (USGS) gauging stations, in concert with estimates of drainage area and basin-averaged precipitation rates¹², allow us to estimate stream power and to compare it with river incision rates. Gauge data indicate that discharge Q scales linearly with the product of drainage area and mean annual precipitation (Methods and Supplementary Fig. 3). Furthermore, measurements of channel width W are consistent with the relationship $W \propto Q^{0.5}$, which agrees with scalings measured in bedrock channels elsewhere (Methods and Supplementary Fig. 4). Substituting $Q \propto AP$ and $W \propto Q^{0.5}$ into $\omega = \rho g QS/W$, we find that gauge data and field measurements on Kaua'i support the following relationship for unit stream power (see Methods): $\omega \propto (AP)^{0.5} S$.

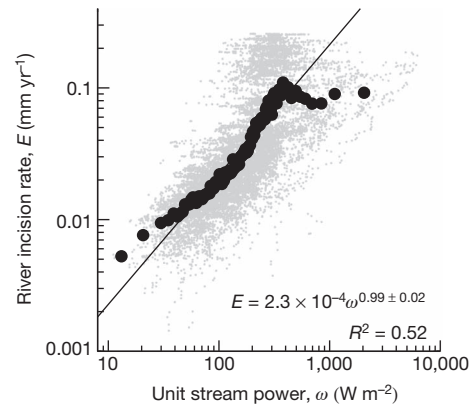


Figure 4 | Dependence of river incision rate on stream power. Million-year-scale river incision rates E (grey points, with black points showing logarithmically binned means) as a function of the stream power per unit bed area $\omega = \rho g QS/W$, calculated using empirical relations $Q = 5.30PA$ and $W = 4.77Q^{0.5}$ (see Methods and Supplementary Figs 3 and 4). River incision rates plateau at the maximum possible long-term average incision rate on Kaua'i (maximum elevation/bedrock age is about $1.6 \text{ km}/4 \text{ Myr} = 0.25 \text{ mm yr}^{-1}$). The regression line is the best power-law fit to all data with $\omega < 379 \text{ W m}^{-2}$ (the stream power at the peak binned incision rate), and is consistent with a linear dependence of river incision rates on stream power below that limit.

We used this relationship to calculate unit stream power ω with an explicit dependence on precipitation rate P at the 13,701 grid points analysed in the time-averaged analysis. Comparing these estimates of ω to our measurements of river incision rate E , we find a linear dependence of million-year-scale river incision rates on stream power (Fig. 4). This result provides rare observational support for the hypothesis that river incision rates are linearly proportional to stream power³. Combined with measurements of discharge, channel width, drainage area and precipitation rate¹², Fig. 4 also offers empirical support for a dependence of river incision rates on the square root of mean annual precipitation: $E \propto \omega \propto (AP)^{0.5} S$ (Methods). The influence of precipitation rate on stream power accounts for a substantial fraction of the variance explained by the linear trend in Fig. 4 (Supplementary Information).

These observations relating river erosional efficiency to precipitation rates on Kaua'i should guide future efforts to include climate explicitly in landscape evolution models. More broadly, the observed correlation between mean annual precipitation and the erosional efficiency of rivers is important because it provides empirical support for a link between Earth's climate and its topographic evolution. This removes a barrier to understanding how topography evolves under spatially variable precipitation—a common feature of Earth's mountain ranges—and it provides a quantitative basis for understanding a critical feedback that governs the long-term development of the continents. Numerical models have long made predictions based on the assumption that precipitation patterns shaped by topography drive spatially variable erosion rates that in turn shape the topography^{2,4,11}, but a strong empirical basis for this feedback has been lacking. It has therefore remained unclear whether, for example, orographic precipitation gradients influence the width, asymmetry and internal structure of mountain ranges⁴. Our measurements across one of Earth's most striking climate gradients demonstrate that such a relationship does exist between precipitation rates and the efficiency of river incision, which suggests that the effects predicted by coupled models of erosion, tectonics and climate may indeed occur.

METHODS SUMMARY

Time-averaged analysis of river incision. We calculated drainage area A by performing steepest descent flow routing on a 10-m digital elevation map. Restricting the analysis to the fluvial network ($A > 10^5 \text{ m}^2$), we calculated S at each grid point as the mean gradient from four grid points upstream to four grid

points downstream. We calculated upstream-averaged mean annual precipitation rates P using a model of mean annual precipitation rates constrained by rain gauge measurements¹². We calculated incision rates by dividing the difference between modern and initial channel elevations by bedrock age²⁷ (Supplementary Fig. 1). We estimated initial channel elevations by projecting modern channels onto palaeosurfaces representing the volcano's initial topography. We reconstructed palaeosurfaces by fitting a thin-plate natural cubic spline across each basin, with spline edges fixed to minimally eroded remnants of the initial topography around the basin's perimeter.

Transient analysis of river incision. We modelled channel profile evolution over the duration of incision, assumed to be the mean local bedrock age. Initial and modern channel profiles were extracted using the same procedure as in the time-averaged analysis. Starting with the initial profile, we used a finite-difference method to solve equation (2) forward in time. We set $n = 1$, consistent with incision rates proportional to stream power³ (Fig. 4), $m = 0.59$ to match the time-averaged analysis, and tested 5,000 values of K . For each channel, we selected the K that minimized the root-mean-square error between the modelled and modern profiles (Fig. 2).

Calculating unit stream power. We calculated the unit stream power as $\omega = \rho g Q S / W$ using empirical relationships for Q as a function of A and P (Supplementary Fig. 3) and W as a function of Q (Supplementary Fig. 4), based on measurements of W and Q at USGS gauging stations on Kaua'i.

Full Methods and any associated references are available in the online version of the paper.

Received 3 September 2012; accepted 5 February 2013.

- Howard, A. D. A detachment-limited model of drainage-basin evolution. *Wat. Resour. Res.* **30**, 2261–2285 (1994).
- Whipple, K. X. Bedrock rivers and the geomorphology of active orogens. *Annu. Rev. Earth Planet. Sci.* **32**, 151–185 (2004).
- Seidl, M. A. & Dietrich, W. E. The problem of channel erosion into bedrock. *Catena*, Suppl. 23, 101–124 (1992).
- Willett, S. D. Orogeny and orography: the effects of erosion on the structure of mountain belts. *J. Geophys. Res. Solid Earth* **104**, 28957–28981 (1999).
- Fernandes, N. F. & Dietrich, W. E. Hillslope evolution by diffusive processes: the timescale for equilibrium adjustments. *Wat. Resour. Res.* **33**, 1307–1318 (1997).
- Riebe, C. S., Kirchner, J. W. & Finkel, R. C. Erosional and climatic effects on long-term chemical weathering rates in granitic landscapes spanning diverse climate regimes. *Earth Planet. Sci. Lett.* **224**, 547–562 (2004).
- Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. Influence of late Cenozoic mountain building on ocean geochemical cycles. *Geology* **16**, 649–653 (1988).
- Snyder, N. P., Whipple, K. X., Tucker, G. E. & Merritts, D. J. Landscape response to tectonic forcing: digital elevation model analysis of stream profiles in the Mendocino triple junction region, northern California. *Geol. Soc. Am. Bull.* **112**, 1250–1263 (2000).
- Stock, J. D. & Montgomery, D. R. Geologic constraints on bedrock river incision using the stream power law. *J. Geophys. Res. Solid Earth* **104**, 4983–4993 (1999).
- Howard, A. D. & Kerby, G. Channel changes in badlands. *Geol. Soc. Am. Bull.* **94**, 739–752 (1983).
- Whipple, K. X. & Tucker, G. E. Dynamics of the stream-power river incision model: implications for height limits of mountain ranges, landscape response timescales, and research needs. *J. Geophys. Res. Solid Earth* **104**, 17661–17674 (1999).
- PRISM Climate Group. <http://prism.oregonstate.edu> (Oregon State University, 2006).
- Poveda, G. & Mesa, O. J. On the existence of Lloro (the rainiest locality on earth): enhanced ocean-land-atmosphere interaction by a low-level jet. *Geophys. Res. Lett.* **27**, 1675–1678 (2000).
- Roe, G. H., Whipple, K. X. & Fletcher, J. K. Feedbacks among climate, erosion, and tectonics in a critical wedge orogen. *Am. J. Sci.* **308**, 815–842 (2008).
- Walling, D. E. & Webb, B. W. in *Background to Hydrogeology* (ed. Gregory, K. J.) 69–100 (Wiley, 1983).
- Burbank, D. W. et al. Decoupling of erosion and precipitation in the Himalayas. *Nature* **426**, 652–655 (2003).
- Dadson, S. J. et al. Links between erosion, runoff variability and seismicity in the Taiwan orogen. *Nature* **426**, 648–651 (2003).
- Reiners, P. W., Ehlers, T. A., Mitchell, S. G. & Montgomery, D. R. Coupled spatial variations in precipitation and long-term erosion rates across the Washington Cascades. *Nature* **426**, 645–647 (2003).
- Moon, S. et al. Climatic control of denudation in the deglaciated landscape of the Washington Cascades. *Nature Geosci.* **4**, 469–473 (2011).
- von Blanckenburg, F. The control mechanisms of erosion and weathering at basin scale from cosmogenic nuclides in river sediment. *Earth Planet. Sci. Lett.* **237**, 462–479 (2005).
- Riebe, C. S., Kirchner, J. W., Granger, D. E. & Finkel, R. C. Minimal climatic control on erosion rates in the Sierra Nevada, California. *Geology* **29**, 447–450 (2001).
- Bookhagen, B. & Strecker, M. R. Spatiotemporal trends in erosion rates across a pronounced rainfall gradient: examples from the southern Central Andes. *Earth Planet. Sci. Lett.* **327–328**, 97–110 (2012).
- Leopold, L. B. & Maddock, T. Jr. *The Hydraulic Geometry of Stream Channels and some Physiographic Implications* (US Geological Survey Professional Paper 252, 1953).
- Montgomery, D. R. & Gran, K. B. Downstream variations in the width of bedrock channels. *Wat. Resour. Res.* **37**, 1841–1846 (2001).
- Dunne, T. & Leopold, L. B. *Water in Environmental Planning* (W. H. Freeman, 1978).
- McDougall, I. Age of shield-building volcanism of Kauai and linear migration of volcanism in the Hawaiian Island chain. *Earth Planet. Sci. Lett.* **46**, 31–42 (1979).
- Sherrod, D. R., Sinton, J. M., Watkins, S. E. & Brunt, K. M. *Geologic Map of the State of Hawai'i Sheet 2 of 8* (US Geological Survey Open-File Report 2007-1089, 2007).
- Garcia, M. O. et al. Petrology, geochemistry and geochronology of Kaua'i lavas over 4.5 Myr: implications for the origin of rejuvenated volcanism and the evolution of the Hawaiian plume. *J. Petrol.* **51**, 1507–1540 (2010).
- Waiawa 943 Climate Summary <http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?hi9253> (Western Regional Climate Center, Desert Research Institute, 2011).
- Mount Waialeale Climate Summary 1047 <http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?hi6565> (Western Regional Climate Center, Desert Research Institute, 2011).
- Chadwick, O. A. et al. The impact of climate on the biogeochemical functioning of volcanic soils. *Chem. Geol.* **202**, 195–223 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by the Massachusetts Institute of Technology. We thank M. Slosberg for assistance with topographic analyses, S. Willett for comments that improved the manuscript and M. Rosener, S. Mukhopadhyay, M. Lamb, B. Mackey, J. Scheingross, J. Stock and C. Blay for field assistance and discussions. We thank the State of Hawaii Agribusiness Development Corporation, Landis Ignacio of the Kekaha Agriculture Association, the State of Hawaii Department of Land and Natural Resources, Divisions of State Parks and of Forestry and Wildlife, the US Fish and Wildlife Service and the Alapai and Napolis families for field access.

Author Contributions K.L.F. and K.L.H. performed the topographic analyses, K.L.H. and J.T.P. conducted the channel evolution modelling, all authors conducted the field work and analysed the data and K.L.F. wrote the paper with input from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.L.F. (ferrier@fas.harvard.edu).

METHODS

Time-averaged analysis of river incision. In the time-averaged analysis, we calculated average river incision rates E at each point along each channel by dividing the depth of river incision $\Delta z = z_{\text{initial}} - z_{\text{modern}}$ by the bedrock age T . We calculated T as the mean age of the lithologic unit into which the channel was incised, with lithologic boundaries obtained from geologic maps²⁷ and bedrock ages determined from published K–Ar analyses^{26,28,32}. Our estimates of mean \pm s.e. bedrock ages for the lithologic units are 1.50 ± 0.12 Myr for the Koloa volcanics^{28,32}; 3.91 ± 0.01 Myr for the Makaweli formation^{28,32}; 3.95 ± 0.05 Myr for the Olokele formation³²; and 4.43 ± 0.45 Myr for the Napali formation²⁶ (Supplementary Fig. 1). The million-year-scale river incision rates calculated in this way are consistent with basin-averaged erosion rates averaged over shorter timescales using independent techniques^{33,34}.

We calculated the depth of river incision by subtracting the river's present elevation (determined from a 10-m digital elevation map³⁵) from the river's initial elevation. We estimated each channel's initial profile by fitting a surface over the channel's catchment and projecting the current channel location upward onto the surface. For each catchment, we approximated the initial surface by fitting a thin-plate smoothing spline to minimally eroded remnants of the initial volcano topography around the catchment perimeter³⁶. In places lacking remnant topography, we fixed spline surfaces to catchment ridgelines. We consider the spline surfaces to be minimum bounds on the elevation of the initial topography, because the elevation of the basin perimeter cannot exceed the elevation of the initial topography. This potentially introduces a systematic error into the estimated erosion rates. However, if this error were significant it would probably make the estimated incision rates more uniform among different channels, because basins that have eroded quickly and have begun to erode the surrounding remnant topography will appear to have artificially slower erosion rates. Thus, if anything, this error would bias our analysis against detecting a relationship between precipitation rates and incision rates. We created spline surfaces with a spline smoothing parameter equal to 1, which corresponds to a natural cubic spline. After extracting initial channel profiles in this manner, we smoothed the profiles as necessary to ensure that channel elevations decreased monotonically from the channel heads to the outlets.

We calculated the channel slope S at each grid point as the mean channel gradient between the fourth grid point upstream and the fourth grid point downstream. We calculated drainage area A with a steepest-descent flow routing algorithm³⁷, and trimmed all channels to include only grid points with $A > 10^5 \text{ m}^2$ to restrict the analysis to the fluvial part of the channel network, as is common in other studies³⁸. We calculated upstream-averaged mean annual precipitation rates P at each grid point using a model of modern mean annual precipitation rates based on measurements from 1971 to 2000, resampled to 10-m resolution^{12,39}.

The 32 channels examined in the time-averaged analysis yielded estimates of A , S and E at 13,701 grid points. We used these values in a multiple linear regression of the log-transformed data to estimate best-fit values for m and n in equation (2) for the entire group of data. That is, we found best-fit values for m and n by fitting $\log_{10}(E)$ against $\log_{10}(K) + m\log_{10}(A) + n\log_{10}(S)$. In Fig. 3a, we report estimates of the river incision coefficient K (calculated at each grid point as $E/(A^m S^n)$) using the global best-fit values of m and n normalized by the mean value of K for all grid points used in the time-averaged analysis ($K_{\text{mean}} = 1.53 \times 10^{-8} \text{ m}^{0.18} \text{ yr}^{-1}$).

Transient analysis of river incision. A limitation of the time-averaged analysis is that it assumes a constant slope through time for each channel point. In the transient analysis, we relaxed this assumption by modelling the transient evolution of the channel profile as it incises into bedrock, permitting slopes to change over time. We conducted the transient analysis on 93 channel segments ranging from 0.43 km to 14.3 km in length. We restricted this analysis to river segments in which the standard deviation of P at each grid point in the upstream drainage area was $< 20\%$ of the mean P , so that each of the analysed river segments had a relatively narrow range of precipitation rates. Upstream average values of P among the selected channels range from 0.63 m yr^{-1} to 8.36 m yr^{-1} . Thus, although each channel segment lies within a relatively narrow band of precipitation rates, the set of channel segments spans a wide range of precipitation rates, and permits exploration of the influence of P on channel incision.

We estimated K for each channel segment by evolving the segment's elevation profile forward in time from its estimated initial profile and identifying the value of K that yielded the best match to the modern profile. We extracted initial channel profiles from estimates of the initial volcanic surface, as in the time-averaged analysis. After extracting initial channel profiles, we resampled each profile to a set of 100 grid points evenly spaced in the along-channel horizontal direction. We then used a finite-difference approximation to equation (2), with slopes calculated with upwind spatial differencing⁴⁰, to evolve the channel forward in time from its initial profile over a time interval equal to the mean age of the bedrock along the channel. Analyses using the minimum and modal bedrock age instead of the mean

bedrock age yielded similar results (Supplementary Information). We imposed a constant lowering rate at the downstream boundary that made the channel segment's downstream boundary match its present elevation at the end of the model run, and allowed the upstream boundary to evolve freely. We used a fixed time step duration Δt of 1,000 years and confirmed that solutions agreed with those obtained with $\Delta t = 500$ years.

We determined drainage areas A at the model grid points by linear interpolation of the channel points extracted from the drainage area map created for the time-averaged analysis. We held these drainage areas constant over the duration of each model run based on the results of an earlier study of Hawaiian channels⁴¹, which found that the relationship between drainage area and along-channel distance did not systematically vary with flow age, suggesting that basin geometry is established early in the evolution of volcanic island topography.

We found the best-fit value of K for each channel segment by modelling the evolution of the segment 5,000 times. In each model run, we used a randomly chosen value for K ranging from 10^{-12} to 10^{-2} (in units of $\text{L}^{-0.18} \text{ T}^{-1}$), a fixed value of $m = 0.59$ (the best-fit value obtained in the time-averaged analysis), and a fixed value of $n = 1$, the slope exponent expected if river incision rates are proportional to stream power³ (Fig. 4). The best-fit value of K was that which minimized the normalized root-mean-square error z_{RMSE} between the modelled channel profile at the end of the model run, z_{model} , and the modern channel profile, z :

$$z_{\text{RMSE}} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (z_{\text{model},i} - z_i)^2}}{\max(z) - \min(z)} \quad (3)$$

The summation is performed over N , the number of grid points along the channel ($N = 100$), the subscript i denotes the i th grid point along the channel profile and $\max(z)$ and $\min(z)$ are the maximum and minimum elevations of the modern channel profile. Figure 3b plots these best-fit values of K , normalized by the mean K for all channel segments used in the transient analysis ($K_{\text{mean}} = 1.45 \times 10^{-7} \text{ m}^{-0.18} \text{ yr}^{-1}$), against the value of P averaged over the drainage basin upstream of the channel's lower boundary.

Calculating unit stream power. In Fig. 4, we plot river incision rates against stream power per unit bed area, calculated as $\omega = \rho g Q S / W$, which describes the energy expended on the bed by steady, uniform flow. We used stream discharge and channel width measurements at USGS gauging stations around Kaua'i⁴² to obtain a relationship for channel width W in terms of discharge Q , and a relationship for Q in terms of drainage area A and upstream-averaged rainfall rates P (ref. 12).

We obtained a relationship for Q by extracting the 90-day recurrence interval discharge from 48 gauging stations with daily discharge measurements. For each station, we calculated the recurrence interval of each daily discharge measurement, $(1 + M)/r$, where M is the number of daily discharge measurements recorded at that station and r is the rank of the discharge measurement, with $r = 1$ representing the largest daily discharge recorded at that station. In the 48 gauging station records that we analysed, M ranged from 438 days to 36,149 days, with a mean $M = 11,888$ days. We linearly interpolated to find the 90-day recurrence interval discharge from the daily discharges and recurrence intervals. We calculated drainage area and upstream basin-averaged mean annual precipitation rates¹² at the sites of the discharge measurements, and then regressed the 90-day Q against the product PA . This revealed a strong linear relationship: $Q = 5.30PA$ ($R^2 = 0.94$; Supplementary Fig. 3), with Q in units of $\text{m}^3 \text{ s}^{-1}$, P in m s^{-1} and A in m^2 . We confirmed that this relationship was not biased by the different time intervals over which discharge measurements were recorded at each station by restricting our analysis to discharge measurements recorded concurrently at a subset of these stations (Supplementary Information). We also examined discharges corresponding to recurrence intervals of 180 days, 1 year, 2 years and 5 years and found comparably strong linear relationships, with progressively larger constants of proportionality for longer recurrence intervals (Supplementary Information).

We determined how channel width W varied with the 90-day recurrence interval discharge Q at the 21 USGS gauging stations that also had channel width measurements. At individual stations, field measurements of channel width and depth as functions of discharge show inflections in channel width and depth at a particular discharge, above which channel width increases more gradually and channel depth increases more rapidly with increasing discharge. Consistent with our field observations of steep-sided trapezoidal channel cross-sections, we interpret this inflection as the discharge at which the streambed is fully submerged and the water surface reaches the base of the channel banks. We chose to analyse the 90-day recurrence interval discharge from the stream gauge records because it is large enough to lie above the inflection in the discharge-width trends for the field measurements at each station, yet small enough to be captured by the field

measurements at all 21 gauging stations (it is near the maximum Q for which channel widths were measured at most stations). Thus, although the 90-day discharge is smaller than the bankfull discharge, the channel width at the 90-day discharge nonetheless appears to record the distance between the channel banks. These measurements yielded a best-fit relationship of $W = 4.3Q^{0.52 \pm 0.09}$ (mean \pm 95% confidence interval; $R^2 = 0.89$; Supplementary Fig. 4), with W in units of m and Q in $\text{m}^3 \text{s}^{-1}$. Because this agrees within uncertainty with a scaling relation of $W \propto Q^{0.5}$ determined from a compilation of width–discharge measurements in bedrock channels elsewhere⁴³, we assumed that the width–discharge measurements on Kaua'i are in fact governed by a square root dependence of W on Q , and we fitted W to the square root of the 90-day recurrence interval Q . This yielded a best-fit relationship of $W = 4.77Q^{0.5}$.

To calculate stream power per unit bed area ω , we substituted the relationships $Q = 5.30PA$ and $W = 4.77Q^{0.5}$ into $\omega = \rho gQS/W$, which yielded the expression $\omega = 0.48\rho g(AP)^{0.5}S$, given the following units for each variable: ρ (kg m^{-3}), g (m s^{-2}), A (m^2), P (m s^{-1}), S (m m^{-1}). We then used estimates of drainage area, channel slope and upstream average mean annual precipitation¹² at the 13,701 grid points analysed in the time-averaged analysis along with values of $\rho = 1,000 \text{ kg m}^{-3}$ and $g = 9.8 \text{ m s}^{-2}$ to calculate unit stream power at each grid point. These estimates of unit stream power are representative of the characteristic 90-day discharge. The 90-day discharge is not necessarily the flow that does the most erosional work, but given that discharge scales linearly with the product of precipitation rates and drainage area for recurrence intervals as long as five years (Supplementary Information), we expect that the stream power at larger discharges would differ only by a constant factor.

32. Clague, D. A. & Dalrymple, G. B. Age and petrology of alkalic postshield and rejuvenated-stage lava from Kauai, Hawaii. *Contrib. Mineral. Petrol.* **99**, 202–218 (1988).
33. Gayer, E., Mukhopadhyay, S. & Meade, B. J. Spatial variability of erosion rates inferred from the frequency distribution of cosmogenic ^3He in olivines from Hawaiian river sediments. *Earth Planet. Sci. Lett.* **266**, 303–315 (2008).
34. Ferrier, K. L. *et al.* Covariation of climate and long-term erosion rates across a steep rainfall gradient on the Hawaiian island of Kaua'i. *Geol. Soc. Am. Bull.* <http://dx.doi.org/10.1130/B30726.1> (in the press).
35. Gesch, D. *et al.* The National Elevation Dataset. *Photogramm. Eng. Remote Sensing* **68**, 5–11 (2002).
36. Brocklehurst, S. H. & Whipple, K. X. Glacial erosion and relief production in the Eastern Sierra Nevada, California. *Geomorphology* **42**, 1–24 (2002).
37. O'Callaghan, J. F. & Mark, D. M. The extraction of drainage networks from digital elevation data. *Computer Vision Graph. Image Process.* **28**, 323–344 (1984).
38. Wobus, C. *et al.* in *Tectonics, Climate, and Landscape Evolution* (eds Willett, S. D., Hovius, N., Brandon, M. T. & Fisher, D. M.) Vol. 398, 55–74 (Geological Society of America Special Papers, 2006).
39. Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L. & Pasteris, P. A knowledge-based approach to the statistical mapping of climate. *Clim. Res.* **22**, 99–113 (2002).
40. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical recipes in C* 2nd edn (Cambridge Univ. Press, 1992).
41. Seidl, M. A., Dietrich, W. E. & Kirchner, J. W. Longitudinal profile development into bedrock—an analysis of Hawaiian channels. *J. Geol.* **102**, 457–474 (1994).
42. US Geological Survey *USGS Water Data for the Nation* <http://waterdata.usgs.gov/nwis> (National Water Information System, 2012).
43. Wohl, E. & David, G. C. L. Consistency of scaling relations among bedrock and alluvial channels. *J. Geophys. Res. Earth Surf.* **113**, <http://dx.doi.org/10.1029/2008jf000989> (2008).

Embryology of Early Jurassic dinosaur from China with evidence of preserved organic remains

Robert R. Reisz¹, Timothy D. Huang², Eric M. Roberts³, ShinRung Peng⁴, Corwin Sullivan⁵, Koen Stein⁶, Aaron R. H. LeBlanc¹, DarBin Shieh⁴, RongSeng Chang⁷, ChengCheng Chiang⁸, Chuanwei Yang⁹ & Shiming Zhong¹⁰

Fossil dinosaur embryos are surprisingly rare, being almost entirely restricted to Upper Cretaceous strata that record the late stages of non-avian dinosaur evolution^{1,2}. Notable exceptions are the oldest known embryos from the Early Jurassic South African sauropodomorph *Massospondylus*^{3,4} and Late Jurassic embryos of a theropod from Portugal⁵. The fact that dinosaur embryos are rare and typically enclosed in eggshells limits their availability for tissue and cellular level investigations of development. Consequently, little is known about growth patterns in dinosaur embryos, even though post-hatching ontogeny has been studied in several taxa⁶. Here we report the discovery of an embryonic dinosaur bone bed from the Lower Jurassic of China, the oldest such occurrence in the fossil record. The embryos are similar in geological age to those of *Massospondylus* and are also assignable to a sauropodomorph dinosaur, probably *Lufengosaurus*⁷. The preservation of numerous disarticulated skeletal elements and eggshells in this monotaxic bone bed, representing different stages of incubation and therefore derived from different nests, provides opportunities for new investigations of dinosaur embryology in a clade noted for gigantism. For example, comparisons among embryonic femora of different sizes and developmental stages reveal a consistently rapid rate of growth throughout development, possibly indicating that short incubation times were characteristic of sauropodomorphs. In addition, asymmetric radial growth of the femoral shaft and rapid expansion of the fourth trochanter suggest that embryonic muscle activation played an important role in the pre-hatching ontogeny of these dinosaurs. This discovery also provides the oldest evidence of *in situ* preservation of complex organic remains in a terrestrial vertebrate.

Monotaxic bone beds are prized by palaeobiologists because they yield large numbers of bones that can reveal patterns of development and growth within a single species⁸. Here we report the discovery of a monotaxic embryonic dinosaur bone bed in Lower Jurassic strata near Dawa, Lufeng County, Yunnan Province, China (specimens housed in the Chuxiong Prefectural Museum, catalogue no. C2019 2A233). The bone bed is equivalent in age to the oldest known dinosaurian embryos, preserved in South Africa³. The Chinese sauropodomorph bone bed is an accumulation of disarticulated skeletal elements representing various stages of embryonic development, rather than a set of articulated skeletons enclosed in eggs⁹.

The bone bed was discovered in the Dark Red Beds or Zhangjia'ao member^{10,11} of the Early Jurassic (Sinemurian, 190–197 Myr ago), Lower Lufeng Formation, roughly 3–5 m below the top of the Formation (Fig. 1). This unit is comparable temporally, environmentally and in faunal content to the Upper Elliot Formation of southern Africa¹² (Supplementary Information section 1), and preserves abundant skeletal remains of basal sauropodomorphs^{13,14}. Taphonomically, the

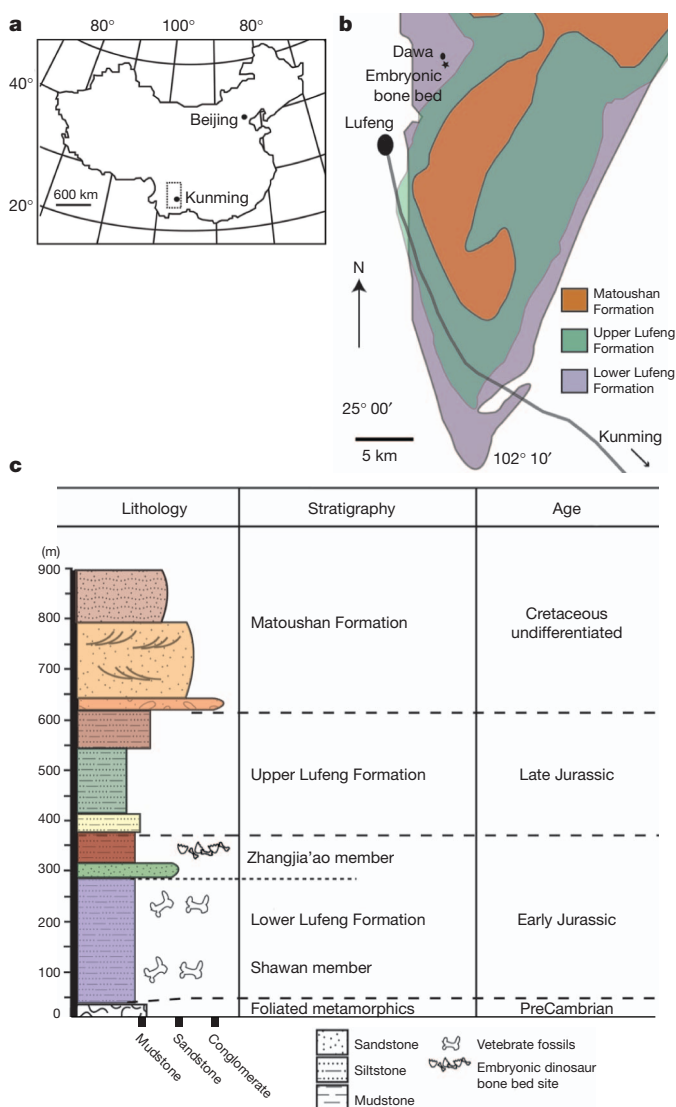


Figure 1 | Location and stratigraphy of Lufeng monotaxic embryonic bone bed. **a**, Map of China with study area in Yunnan Province shown by inset box. **b**, General geological map of the study area. **c**, Stratigraphic section showing location of the embryonic bone bed within Zhangjia'ao member of Lower Lufeng Formation.

¹Department of Biology, University of Toronto Mississauga, Mississauga, Ontario L5L 1C6, Canada. ²National Chung Hsing University, Taichung 402, Taiwan. ³School of Earth and Environmental Sciences, James Cook University, Townsville, Queensland 4811, Australia. ⁴Medical College Institute of Oral Medicine, National Cheng Kung University, Tainan 701, Taiwan. ⁵Key Laboratory of Evolutionary Systematics of Vertebrates, Institute of Vertebrate Paleontology and Paleoanthropology, 100044 Beijing, China. ⁶Steinmann Institut für Geologie, Mineralogie und Paläontologie, University of Bonn, 53115 Bonn, Germany. ⁷Department of Optics and Photonics, National Central University, Chung-Li 32001, Taiwan. ⁸National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan. ⁹Lufeng County Dinosaur Museum, Lufeng, Yunnan, China. ¹⁰Chuxiong Prefectural Museum, 675000 Chuxiong, Yunnan, China.

10–20-cm-thick bone bed is characterized by the presence of completely disarticulated skeletal elements at various stages of embryonic development (Fig. 2), with calcium carbonate nodules often surrounding tightly packed appendicular skeletal elements. One nodule contains a high concentration of eggshell fragments that were apparently derived from soil compaction of a single egg, making it possible to obtain microstructural information about the oldest known terrestrial

vertebrate eggshell (Supplementary Information section 2 and Supplementary Figs 2.1 and 2.2).

We interpret the bone bed as a para-autochthonous assemblage, formed by low-energy flooding and slow inundation of a colonial nesting site. The host sediment is a heavily bioturbated, massive siltstone, throughout which are dispersed isolated skeletal elements, eggshell fragments and the small, fossil-rich nodules of calcium carbonate.

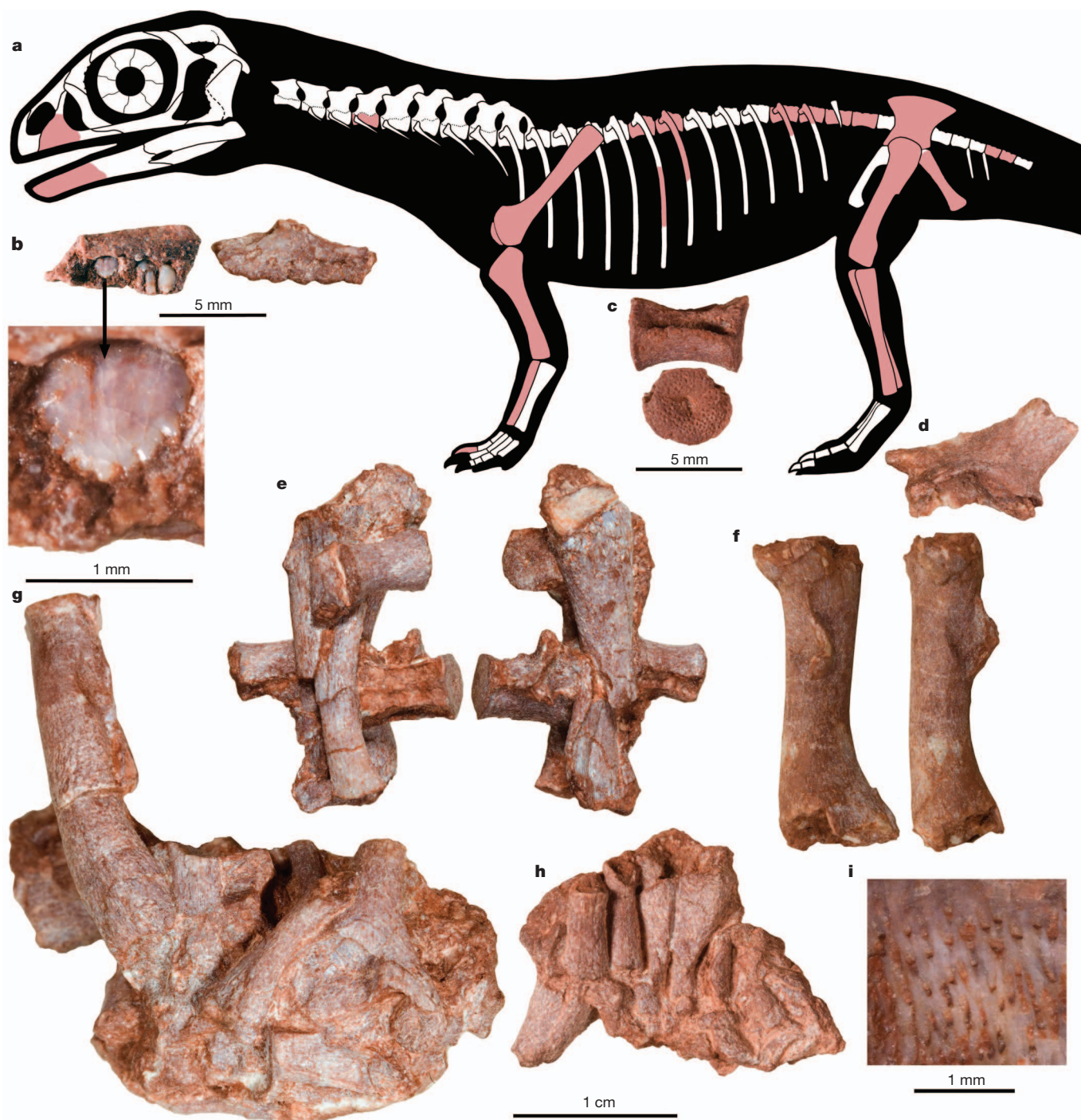


Figure 2 | Sauropodomorph dinosaur embryonic skeletal elements from the Lufeng bone bed. **a**, Reconstructed embryonic skeleton of Early Jurassic sauropodomorph (using *Massospondylus* as a model, not to scale), showing in dark red the known elements from Dawa (Chuxiong Prefectural Museum, no. C2019 2A233). Exact skeletal positions of numerous centra, ribs and distal limb elements are difficult to determine. **b**, Left maxillae in ventromedial and labial views, respectively, with enlarged view of partially erupted tooth. **c**, Mid-dorsal centrum in lateral and anterior views. **d**, Left ilium in lateral view. **e**, Right

scapula, vertebrae and left humerus preserved in one nodule. **f**, Right femur in posterolateral and medial views, showing prominence and shape of the fourth trochanter. **g**, Large right femur preserved with ribs and various other skeletal elements in nodule. **h**, Embryonic limb elements and ribs showing alignment along long axes. **i**, Close-up of proximal end of right tibia, showing external foramina of primary cavities (vascular canals). Scale bar, 1 cm, unless otherwise shown.

There are no preserved nest structures or uncrushed eggs. The lack of coarse-grained sediment, coupled with the apparent sorting and concentration of bones (Fig. 2c, f, g) that are at various developmental stages, and hence from different nests, indicates that the bone bed is not an *in situ* nest or catastrophic death assemblage. We believe that inundation and ponding, followed by weak currents and simple wave action, explain the hydrodynamic sorting and non-random orientation of the disarticulated embryonic elements. Transport must have been minimal given the high preservational quality of the delicate, poorly ossified embryonic bones and <100- μ m-thick eggshell. The embryonic bones and eggshell fragments were eventually buried and subjected to pedogenic processes, including precipitation of carbonate nodules that encase the bones.

The sample of more than 200 bones includes dozens of isolated cervical, dorsal, and caudal centra, rib fragments, femora and other limb elements, scapulae, an ilium, and a few skull elements (Fig. 2). These specimens are less ontogenetically advanced in multiple respects than some previously known sauropodomorph, theropod and ornithischian embryonic skeletons that were discovered inside intact eggs, demonstrating that the Lufeng specimens represent embryos rather than hatchlings¹⁵. Conspicuously embryonic features include the presence of teeth that do not protrude beyond the alveolar edges of the maxilla and dentary, centra with large notochordal canals and deeply pitted articular surfaces, and the universal presence of extensive primary vascular spaces that are open to the surface^{16,17} (Supplementary Information section 3 and Supplementary Fig. 3).

The embryonic bones were compared with previously known saurischian and ornithischian embryos, and found to share detailed resemblances with embryos of other sauropodomorphs but not with those of ornithischians or theropods^{1–3}. In particular, identification of the Lufeng specimens as sauropodomorph was greatly facilitated by their similarity to the well known, articulated *Massospondylus* embryos¹⁸. Interpretation of the embryonic bones as representing a basal sauropodomorph is based on numerous features that are synapomorphies at various levels within basal Sauropodomorpha, and on a phylogenetic analysis using data from a recent study¹⁹. This analysis places the specimens well within the sauropodomorph clade but well outside Sauropoda, and supports their tentative referral to the well known Lufeng Formation sauropodomorph *Lufengosaurus*. Within Sauropodomorpha, the maxilla and its dentition show specific morphological resemblances to *Lufengosaurus*⁷. However, two other basal sauropodomorphs have also been recovered from the Lower Lufeng Formation of Yunnan¹⁴, making identification of embryonic specimens as *Lufengosaurus* inescapably tentative (Supplementary Information section 4 and Supplementary Figs 4.1 and 4.2).

Histological study of the Lufeng embryonic specimens provides an unprecedented window into embryonic growth in a dinosaur because these fossils represent numerous individuals at various stages of embryonic development. For example, three thin-sectioned dorsal vertebrae show different stages in the embryonic development of the notochordal canal¹⁶ (Fig. 3), a feature that is absent in individuals beyond the hatchling stage. Longitudinal sections of two vertebrae (Fig. 3a, b) show that the cranial and caudal ends consist mostly of hypertrophied calcified cartilage. The mid-regions of the vertebrae show an initial stage of highly cancellous bone deposition with numerous primary cavities ('vascular spaces'), indicative of very fast growth²⁰, and there is no evidence of any bone remodelling. A transverse section through the third vertebra (Fig. 3c) shows the notochordal canal as a large tunnel through the middle of the centrum, and reveals erosion cavities that indicate resorption of the cartilaginous precursor.

The sample includes 24 femora ranging from 2.6 to 4.5 mm in mid-shaft diameter, and 12 to 22 mm in length. They represent individuals from multiple nests¹⁸, permitting the first morphometric analysis of embryonic growth in a dinosaur (Fig. 4a, Supplementary Information section 5 and Supplementary Figs 5.1–5.4). Thin sections through the mid-shaft regions and fourth trochanters of three femora of different

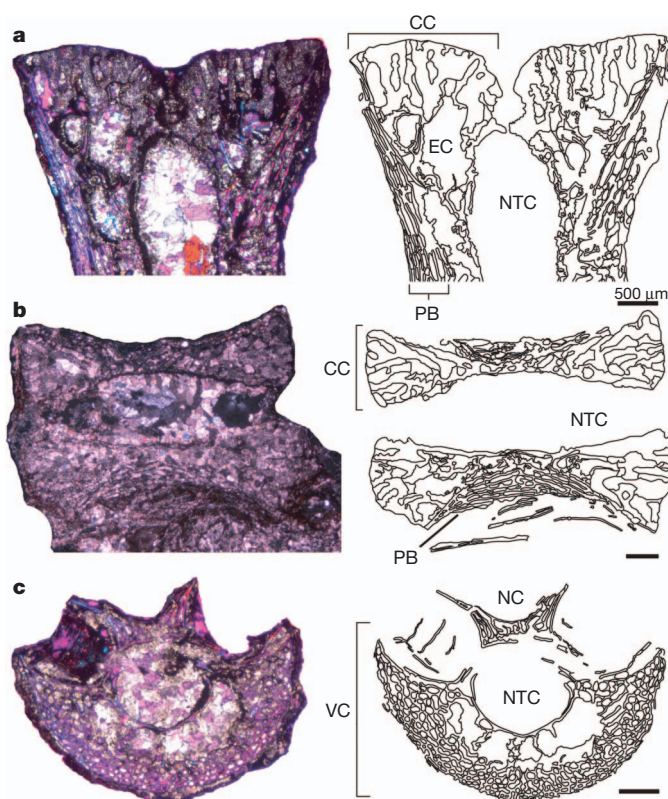


Figure 3 | Embryonic vertebral histology (C2019 2A233). **a**, Largest dorsal centrum, longitudinal section, cranial portion showing initial closure of notochordal canal and presence of erosion cavities with endochondral bone in the calcified cartilage. **b**, Smallest dorsal centrum, longitudinal section of whole bone, representing earlier embryonic stage with widely open notochordal canal. **c**, Intermediate-sized dorsal centrum, transverse section, showing notochordal canal. All scale bars represent 500 μ m. Photographs taken with cross polar light with a lambda filter. CC, calcified cartilage; EC, erosion cavities; NC, neural canal; NTC, notochordal canal; PB, periosteal bone; and VC, vertebral centrum.

diameters also illustrate the development and ossification of the femur (Fig. 4b–d). The cross-sections show major differences in periosteal bone distribution, orientation of the vascular spaces (primary cavities), and size and level of ossification of the fourth trochanter (insertion site of the primary propulsive muscle of the hindlimb). In the smallest, least ossified femur (Fig. 4b), the fourth trochanter is small. In the mid-sized femur (Fig. 4c) the trochanter is more prominent, but contains little woven bone tissue. In the largest femur (Fig. 4d), the fourth trochanter is large and fully ossified. Mid-shaft cross-sections of seven femora, including the three already listed (Fig. 4b–d) show significantly greater vascularity (ratio of primary cavity area to total cross-sectional area of the cortex ranging from 56% to 65%) than in other dinosaurs, indicating a sustained very rapid rate of growth²⁰. In addition, the femoral medullary cavity increases in diameter throughout embryonic development (Supplementary Fig. 5.4), indicating that, although the embryonic femur is composed entirely of primary bone, reshaping by endosteal bone resorption in the medullary cavity occurs even at this early ontogenetic stage. The high level of vascularity is the first known evidence that sauropodomorph embryos probably grew faster than those of extant birds and other dinosaurs, a circumstance that may imply that sauropodomorphs had shorter incubation times than their contemporaries. This capacity for rapid growth was evidently maintained after hatching²¹, explaining the ability of sauropodomorphs to consistently achieve larger adult size than their dinosaurian contemporaries and in some cases reach gigantic proportions.

Extant vertebrates can display considerable limb and body movement before birth or hatching, involving muscle activation that mediates

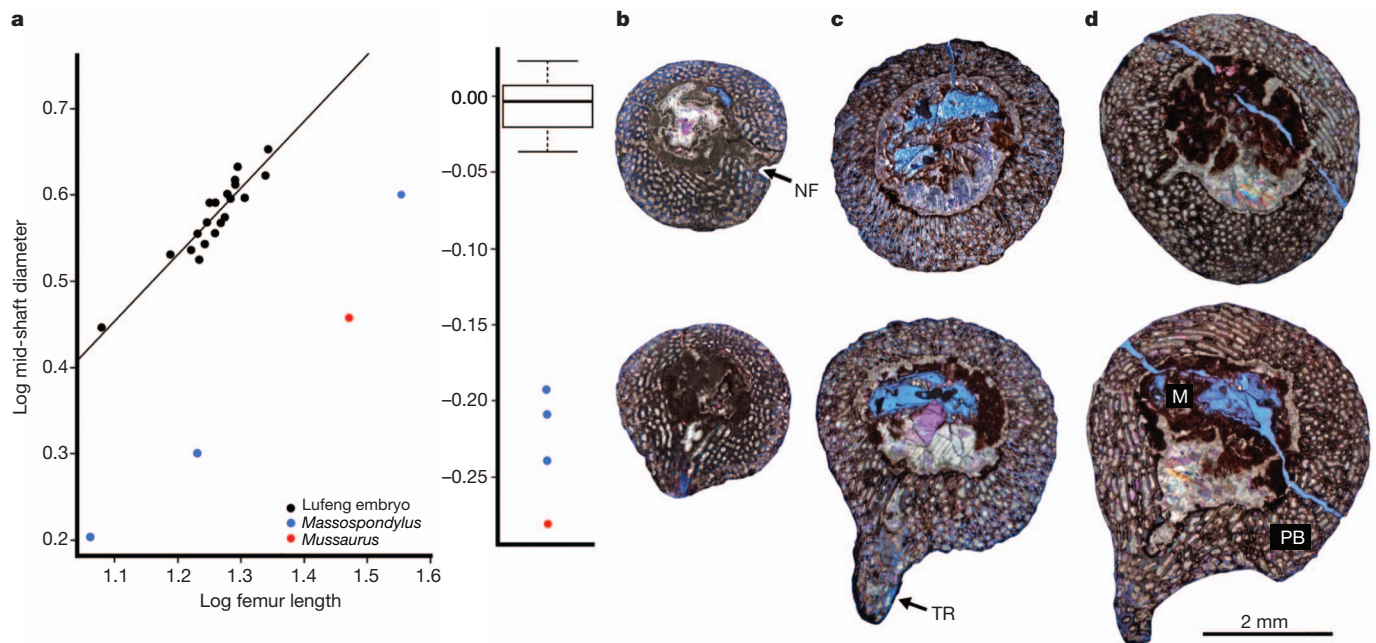


Figure 4 | Embryonic femoral morphometric analysis and histology (C2019 2A233). **a**, Results of regression analysis of Lufeng embryonic femora (black dots) showing growth trajectory of femoral mid-shaft diameter relative to length of femur, and box-plot of residuals. Lufeng femora show strong correlation between femoral length and shaft diameter ($R^2 = 0.8713$) and there are no outliers in this sample (see Supplementary Information section 5 and Supplementary Figs 5.1–5.4 for data and additional morphometric analyses). Two embryonic femora and one hatchling femur of the sauropodomorph

Massospondylus (blue dots), and a hatchling femur of the sauropodomorph *Mussaurus* (red dot), are well outside the range of variation of the Lufeng embryos. **b–d**, Smallest to largest femora, sectioned transversely at mid-diaphysis above, and at the level of the fourth trochanter below. Photographs taken with cross polar light with a lambda filter. M, medullary cavity; NF, nutrient foramen; PB, primary periosteal bone with primary cavities ('vascular canals'); TR, fourth trochanter.

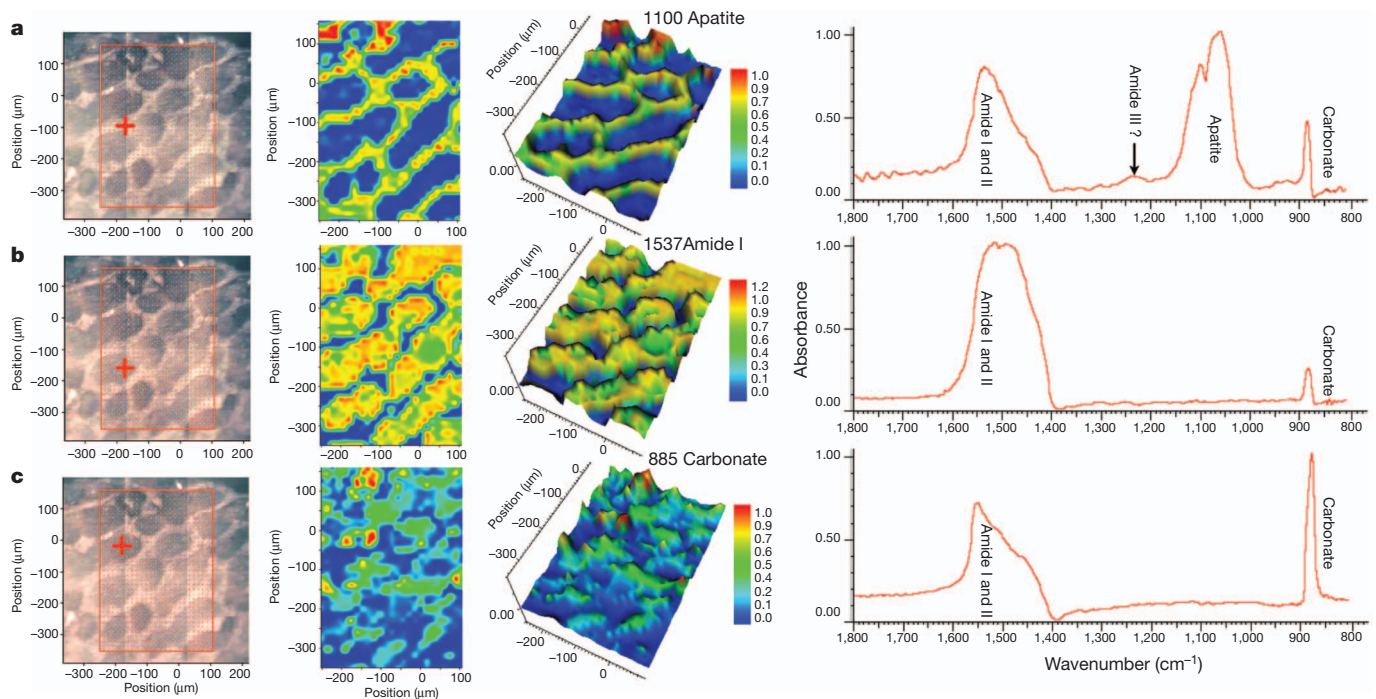


Figure 5 | Synchrotron radiation Fourier transform infrared spectroscopy analysis of embryonic femur. **a–c**, Different areas targeted within the bone. Images in each row include, from left to right: mosaic composed of 12 individual optical images ($150 \times 180 \mu\text{m}$ each), showing total Fourier transform infrared (FTIR) scanned area (red box), specific point targeted for analysis (red cross) and $15 \mu\text{m}$ step size (red dots); two-dimensional and three-dimensional FTIR distributions of absorption for the spectral band showing the highest intensity at the targeted point, with blue and red corresponding to low and high absorption, respectively; and Kramers–Kronig-transformed infrared

spectrum for the targeted point. **a**, Primary bone tissue showing an apatite peak within primary bone but not within vascular spaces; primary bone tissue also shows an amide peak at $1,500\text{--}1,700 \text{ cm}^{-1}$, within apatite crystal. **b**, Edge of vascular space, near primary bone tissue, showing that margins of vascular spaces are characterized by the $1,537 \text{ cm}^{-1}$ amide peak but lack the apatite peak. **c**, Central area of a vascular canal, showing a 885 cm^{-1} carbonate peak within vascular spaces; a modest amide peak at $1,537 \text{ cm}^{-1}$ is also present. We interpret the carbonate peak as the result of calcite infilling of the vascular canals.

skeletal development^{22–24}. In mice and chickens this epigenetic phenomenon results in differential (asymmetrical) thickening of the walls of the long bones for improved load-bearing, resulting in geometry similar to that seen in the embryonic sauropodomorph femora. Similarly, sustained growth of skeletal crests and flanges depends on activation of the muscles attached to them, an observation that should be applicable to the dinosaurian fourth trochanter. It is probable that the uneven thickening of the femoral walls in the Lufeng specimens, the circumferential orientation of the primary vascular cavities, and the growth of the fourth trochanter (Fig. 4b–d) all depended upon muscle contraction and embryonic motility, an important mechanism for building a skeleton capable of coping with the functional demands encountered by the neonate. This discovery adds the first fossil evidence of such epigenetic phenomena to a growing body of research that documents their significance in extant model organisms.

The embryonic bones were also studied using synchrotron radiation-Fourier transform infrared (SR-FTIR) spectroscopy²⁵. In contrast to previous studies of organic residues based on extracts obtained by decalcifying samples of bone, our approach targeted particular tissues *in situ* (Fig. 5). This made it possible to detect the preservation of organic residues, probably direct products of the decay of complex proteins, within both the fast-growing embryonic bone tissue and the margins of the vascular spaces (Fig. 5a, b). This is indicated by the multiple amide peaks revealed by both infrared (1,500–1,700 cm⁻¹ strong band from amide I and II, and 1,200–1,300 cm⁻¹ weak band from amide III) and Raman spectroscopy (amide A peak at 3,264 cm⁻¹) (Supplementary Figs 6.1 and 6.2). Previous reports of preserved dinosaur organic compounds, or ‘dinosaurian soft tissues’^{26–28}, have been controversial because it was difficult to rule out bacterial biofilms or some other form of contamination as a possible source of the organics²⁹. Our results clearly indicate the presence of both apatite and amide peaks within woven embryonic bone tissue (Fig. 5a), which should not be susceptible to microbial contamination or other post-mortem artefacts.

METHODS SUMMARY

Fossil preparation was done manually under a dissecting microscope. Thin sections ranging from 30 to 50 µm in thickness were photographed using a Nikon AZ100 microscope with a lambda filter. Image J software was used to calculate the percentage vascularity (ratio between the total area of the primary cavities and the overall area of the cortex) of each femoral thin section. NIS Elements imaging software for the Nikon DS-Fi2 mounted digital camera was used both for photography and confirmation of percentage vascularity. For FTIR analysis, infrared spectral line scans and mapping data were collected using the SR-FTIR spectro-microscopy facility at the National Synchrotron Radiation Research Center (NSRRC) beamline 14A1 (BL14A1) in Taiwan. The spectra were recorded in reflectance mode from each sample section using a Thermo Nicolet 6700 FTIR spectrometer and continuum infrared microscope with the following settings: resolution 4 cm⁻¹, step size 15 µm, aperture size 30 µm and 128 scans. Peak position and baseline corrections were performed using OMNIC peak resolving software.

Received 25 October 2012; accepted 4 February 2013.

1. Carpenter, K., Hirsch, K. F. & Horner, J. R. (eds). *Dinosaur Eggs and Babies* (Cambridge Univ. Press, 1994).
2. Chiappe, L. M. *et al.* Sauropod dinosaur embryos from the Late Cretaceous of Patagonia. *Nature* **396**, 258–261 (1998).
3. Reisz, R. R., Scott, D., Sues, H.-D., Evans, D. C. & Raath, M. A. Embryos of an Early Jurassic prosauropod dinosaur and their evolutionary significance. *Science* **309**, 761–764 (2005).
4. Reisz, R. R., Evans, D. C., Roberts, E. M., Sues, H.-D. & Yates, A. M. Oldest known dinosaurian nesting site and reproductive biology of the Early Jurassic sauropodomorph *Massospondylus*. *Proc. Natl Acad. Sci. USA* **109**, 2428–2433 (2012).
5. de Ricqlès, A., Mateus, O., Antunes, M. T. & Taquet, P. Histomorphogenesis of embryos of Upper Jurassic theropods from Lourinhã (Portugal). *C. R. Acad. Sci. IIA* **332**, 647–656 (2001).
6. Padian, K., de Ricqlès, A. J. & Horner, J. R. Dinosaurian growth rates and bird origins. *Nature* **412**, 405–408 (2001).

7. Young, C. C. A complete osteology of *Lufengosaurus huenei* Young (gen. et sp. nov.) from Lufeng, Yunnan, China. *Palaeontol. Sinica New Series C* **7**, 1–53 (1941).
8. Rogers, R. R., Eberth, D. A. & Fiorillo, A. R. (eds). *Bonebeds, Genesis, Analysis, and Paleobiological Significance* (Chicago Univ. Press, 2007).
9. Carpenter, K. *Eggs, Nests, and Baby Dinosaurs. A Look at Dinosaur Reproduction* (Indiana Univ. Press, 1999).
10. Bien, M. N. “Red Beds” of Yunnan. *Bull. Geol. Soc. China* **21**, 159–198 (1941).
11. Fang, X. *et al.* in *Proc. Third National Stratigraphical Congress of China*. 208–214 (Geological Publishing House, 2000).
12. Weishampel, D. B., Dodson, P. & Osmólska, H. (eds). *The Dinosauria* (Univ. California Press, 2004).
13. Sun, A. G. & Cui, K. H. in *The Beginning of the Age of Dinosaurs* (ed. Padian, K.) 275–278 (Cambridge Univ. Press, 1986).
14. Galton, P. M. & Upchurch, P. in *The Dinosauria* (eds Weishampel D. B., Dodson, P. & Osmólska, H.) 232–258 (Univ. California Press, 2004).
15. Kundrát, M., Cruickshank, A. R. I., Manning, T. W. & Nudds, J. Embryos of therizinosaurid theropods from the Upper Cretaceous of China: diagnosis and analysis of ossification patterns. *Acta Zool. (Stockh.)* **89**, 231–251 (2008).
16. Fleming, A., Keynes, R. J. & Tannahill, D. The role of the notochord in vertebral column formation. *J. Anat.* **199**, 177–180 (2001).
17. Francillon-Vieillot, H. *et al.* in *Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends* (ed. Carter, J. G.) 471–530 (Van Nostrand Reinhold, 1990).
18. Reisz, R. R., Sues, H. D., Evans, D. C. & Scott, D. Embryonic skeletal anatomy of the sauropodomorph dinosaur *Massospondylus* from the Lower Jurassic of South Africa. *J. Vertebr. Palaeontol.* **30**, 1653–1665 (2010).
19. Apaldetti, C., Pol, D. & Yates, A. The postcranial anatomy of *Coloradisauros brevis* (Dinosauria: Sauropodomorpha) from the Late Triassic of Argentina and its phylogenetic implications. *Palaeontology* <http://dx.doi.org/10.1111/j.1475-4983.2012.01198.x> (22 November 2012).
20. Horner, J. R., Padian, K. & Ricqlès, A. D. Comparative osteohistology of some embryonic and perinatal archosaurs: developmental and behavioural implications for dinosaurs. *Paleobiology* **27**, 39–58 (2001).
21. Sander, M. P. Longbone histology of the Tendaguru sauropods: implications for growth and biology. *Paleobiology* **26**, 466–488 (2000).
22. Müller, G. B. Embryonic motility: environmental influences and evolutionary innovation. *Evol. Dev.* **5**, 56–60 (2003).
23. Sharir, A., Stern, T., Rot, C., Shahar, R. & Zelzer, E. Muscle force regulates bone shaping for optimal load-bearing capacity during embryogenesis. *Development* **138**, 3247–3259 (2011).
24. Blizt, E. *et al.* Bone ridge patterning during musculoskeletal assembly is mediated through SCX regulation of *Bmp4* at the tendon-skeleton junction. *Dev. Cell* **17**, 861–873 (2009).
25. Kong, J. & Shaoning, Y. Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochim. Biophys. Sin. (Shanghai)* **39**, 549–559 (2007).
26. Schweitzer, M. H., Wittmeyer, J. L., Horner, R. H. & Toporski, J. K. Soft-tissue vessels and cellular preservation in *Tyrannosaurus rex*. *Science* **307**, 1952–1955 (2005).
27. Schweitzer, M. H. *et al.* Analyses of soft tissue from *Tyrannosaurus rex* suggest the presence of protein. *Science* **316**, 277–280 (2007).
28. Lindgren, J. *et al.* Microspectroscopic evidence of Cretaceous bone proteins. *PLoS ONE* **6**, e19445 (2011).
29. Kaye, T. G., Gaugler, G. & Sawlowicz, Z. Dinosaurian soft tissue interpreted as bacterial biofilms. *PLoS ONE* **3**, e2808 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Scott for specimen preparation, photography, and figure preparation; N. Campione for morphometric analysis; C. Apaldetti for the data matrix; O. Dülfer for thin sections; G. Grellet-Tinner, M. Sander, J. Steigler, P. Barrett and E. Prondvai for discussion; C. Chu and X. J. Lin for research support; S. P. Modesto and C. Brown for field assistance; J. Liu for assistance in Lufeng; and C. C. Wang, Y. F. Song, Y. C. Lee and H. S. Sheu for help with various experiments at the National Synchrotron Radiation Research Center, Taiwan. Research support was provided by NSERC Discovery and SRO Grants (Canada), University of Toronto, DFG FOR 533 (contribution 130) (Germany), NSC 100-2116-M-008-016 (Taiwan), Ministry of Education (Taiwan) under the NCKU Aim for the Top University Project, Chinese Academy of Sciences and National Natural Science Foundation of China (41150110341).

Author Contributions R.R.R. jointly conceived and designed the project with T.D.H.; R.R.R. wrote the paper, and supervised preparation and scientific illustration of specimens; T.D.H., E.M.R., C.S., K.S. and A.R.H.L., contributed to the manuscript; R.R.R., T.D.H., E.M.R., C.S., R.C. and C.Y. contributed to field work; T.D.H., S.P. and D.S. supervised and completed multimodal optical and chemical spectroscopic analyses; K.S., A.R.H.L. and C.C. prepared slides and illustrated thin sections; R.R.R., T.D.H., K.S., E.M.R., S.P. and C.S. wrote the Supplementary Information; T.D.H., R.C., C.Y. and S.Z. provided logistical support for field work and research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.R.R. (robert.reisz@utoronto.ca).

Predominant archaea in marine sediments degrade detrital proteins

Karen G. Lloyd^{1,2*}, Lars Schreiber^{1*}, Dorthe G. Petersen¹, Kasper U. Kjeldsen¹, Mark A. Lever¹, Andrew D. Steen², Ramunas Stepanauskas³, Michael Richter⁴, Sara Kleindienst⁵, Sabine Lenk⁵, Andreas Schramm¹ & Bo Barker Jørgensen¹

Half of the microbial cells in the Earth's oceans are found in sediments¹. Many of these cells are members of the Archaea², single-celled prokaryotes in a domain of life separate from Bacteria and Eukaryota. However, most of these archaea lack cultured representatives, leaving their physiologies and placement on the tree of life uncertain. Here we show that the uncultured miscellaneous crenarchaeotal group (MCG) and marine benthic group-D (MBG-D) are among the most numerous archaea in the marine sub-sea floor. Single-cell genomic sequencing of one cell of MCG and three cells of MBG-D indicated that they form new branches basal to the archaeal phyla Thaumarchaeota³ and Aigarchaeota⁴, for MCG, and the order Thermoplasmatales, for MBG-D. All four cells encoded extracellular protein-degrading enzymes such as gingipain and clostripain that are known to be effective in environments chemically similar to marine sediments. Furthermore, we found these two types of peptidase to be abundant and active in marine sediments, indicating that uncultured archaea may have a previously undiscovered role in protein remineralization in anoxic marine sediments.

In the cold anoxic sediments underlying most of the Earth's oceans, the only metabolisms known for cultured archaea are methane production from simple carbon substrates, and methane consumption⁵. Recent isotopic evidence, however, has shown that sedimentary archaea can be heterotrophic⁶, but potential carbon substrates remain unknown. Intriguingly, detrital proteins are the largest components of

marine organic matter that can be chemically characterized⁷, and they are degraded slowly by thus-far-unknown microbes in anoxic sediments⁸. If heterotrophic archaea were able to degrade proteins, such a finding would change our basic conception of the marine sedimentary carbon cycle, as it is generally assumed that bacteria drive the primary remineralization of complex organic matter.

Globally, marine subsurface archaea are often dominated by members of the MCG and MBG-D⁹ (Fig. 1). We analysed these groups of archaea present in Aarhus Bay, Denmark (Supplementary Fig. 1), in organic-rich marine sediments where microbial activity is high at the surface but drops to low values similar to those found in deep oceanic sediments a few metres below the sea floor¹⁰. Here, MCG and MBG-D are abundant, based on 16S ribosomal RNA gene polymerase chain reaction (PCR) amplicon sequence libraries and quantitative PCR (Supplementary Fig. 2 and Supplementary Tables 1–3). Their abundance is independent of the major biogeochemical zones of sulphate reduction and methanogenesis, as has been noted previously for MCG⁹, suggesting that they have different metabolic pathways from these types of anaerobic respiration.

However, the taxonomic marker 16S rRNA gene cannot be used to infer physiologies for MCG and MBG-D because no member or close relative of these groups has ever been grown in laboratory culture. We therefore obtained large genomic sequences from single cells to couple taxonomic genes to those encoding other cellular functions^{11,12}. We used density gradient centrifugation to extract intact cells from a

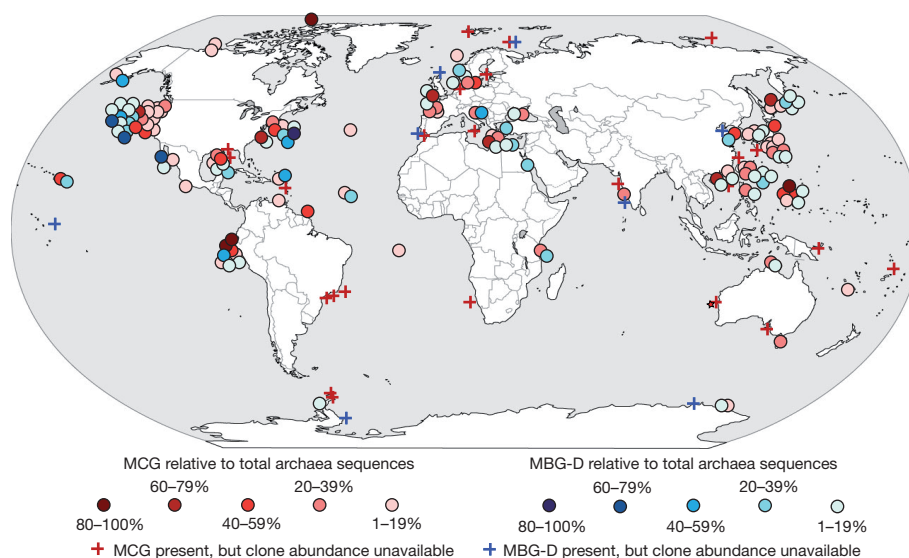


Figure 1 | Global marine occurrence of miscellaneous crenarchaeotal group (MCG) and marine benthic group D (MBG-D). Relative abundance of 16S rRNA gene sequences in clone libraries from marine sediments for MCG (red)

and MBG-D (blue). For some (crosses), sequence abundance information was unavailable.

¹Center for Geomicrobiology, Department of Bioscience, Aarhus University, Aarhus 8000, Denmark. ²University of Tennessee, Knoxville, Tennessee 37996, USA. ³Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine 04544, USA. ⁴Ribocon GmbH, Bremen 28359, Germany. ⁵Max Planck Institute for Marine Microbiology, Bremen 28359, Germany.

*These authors contributed equally to this work.

sediment depth of 10 cm, sorted individual cells with fluorescence-activated cell sorting (FACS), amplified their genomic DNA with multiple displacement amplification (MDA), and screened for taxonomic identification via 16S rRNA gene sequences (Methods and Supplementary Figs 3 and 4). We selected a genome amplified from a single MCG cell (MCG_E09, which has NCBI name *Thaumarchaeota* archaeon SCGC AB-539-E09) and three genomes from single MBG-D cells (MBG-D_N05, MBG-D_F20 and MBG-D_C06, which have NCBI names *Thermoplasmatales* archaeon SCGC AB-539-N05, SCGC AB-539-F20 and SCGC AB-539-C06 in the NCBI database) for whole-genome sequencing (Supplementary Tables 4–6). Total *de novo* assembly sizes for the single amplified genomes (SAGs) ranged from 0.593 to 1.037 megabases (Mb) with 73–172 contigs. Quality control showed no evidence of contamination (Methods and Supplementary Fig. 5). Estimated genome coverage was 32–70% (Table 1), due to uneven or biased genomic amplification during MDA¹¹.

Phylogenetic analyses based on a concatenation of genes conserved in single copies in all archaea in the Integrated Microbial Genomes (IMG) database (Supplementary Fig. 6 and Supplementary Table 7) placed the MCG cell on a deep branch within the phyla *Thaumarchaeota*³ and *Aigarchaeota*⁴, and the MBG-D cells basal to or just inside group MG-II in the order *Thermoplasmatales* in the phylum *Euryarchaeota* (Fig. 2 and Supplementary Fig. 7). The MCG-E09 placement agrees with phylogenies based on single taxonomic marker genes that show that MCG are distinct from the *Crenarchaeota*¹³. However, the three MBG-D SAGs are not monophyletic relative to MG-II, the only other uncultured group of *Thermoplasmatales* with genomic information. This suggests either that partial complements of archaeal conserved genes cannot resolve fine-scale phylogenies or that the evolutionary history of the MBG-D is more complex than that derived from 16S rRNA genes. The 16S rRNA gene sequences from the single cells are >98% similar to other environmental sequences, making the cells representatives of these uncultured groups (Supplementary Fig. 8).

All four single cells contained predicted extracellular cysteine peptidases found in anaerobic protein-degrading bacteria: clostripain (Merops family C11, 1 copy in MCG_E09 and 2 copies in MBG-D_N05), gingipain (Merops family C25, 4, 12 and 6 copies in MBG-D_N05, MBG-D_F20 and MBG-D_C06, respectively), papain (Merops family C1A, 2 and 1 copies in MBG-D_N05 and MBG-D_F20, respectively) and pyroglutamyl peptidase (Merops family C15, 1 copy in MBG-D_F20) (Fig. 3 and Supplementary Tables 8 and 9)¹⁴. Clostripain and gingipain are secreted endopeptidases that are specific for arginine (or sometimes lysine, for gingipain) in the primary amino acid position of the substrate¹⁵. Papain and pyroglutamyl peptidase are often cytosolic or lysosomal; the first cleaves a wide variety of substrates and the second removes a pyroglutamate residue from the amino terminus of a peptide¹⁵. Each SAG gene homologue encodes a complete active site and signal sequences for extracellular transport; some may be coexpressed as they cluster on the genome (Supplementary Table 9). The clostripain from MCG_E09 is closely

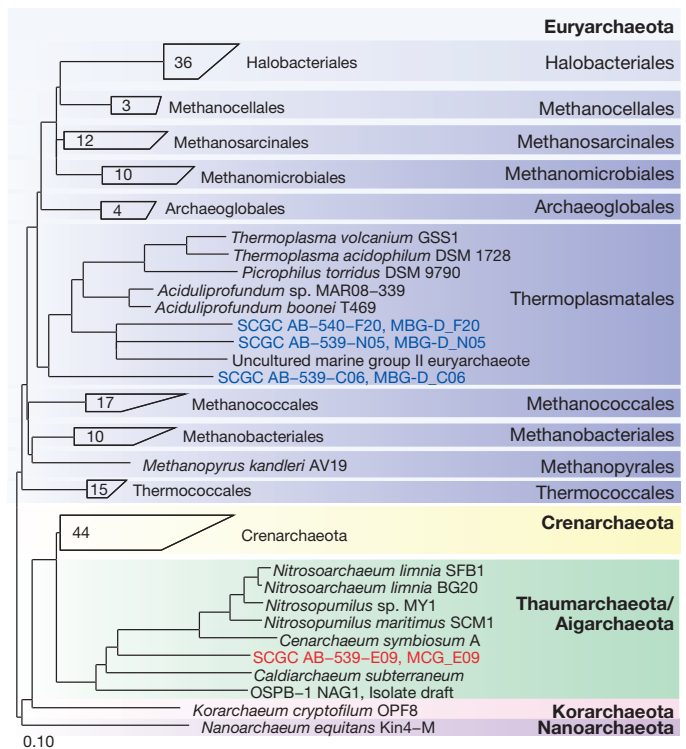


Figure 2 | Evolutionary placement of SAGs. Consensus of maximum likelihood (RAxML) trees of concatenated core archaeal conserved single-copy genes (individual trees shown in Supplementary Fig. 7). Phyla (bold) and orders are labelled. Numbers of genomes in collapsed clades are written on the boxes.

related to that of *Clostridia* spp. (Supplementary Fig. 9), and contains binding sites for the cofactor Ca^{2+} (Fig. 3d)¹⁶. In MBG-D_N05, clostripain is closely related to that of *Aciduliprofundum boonei*, a hyperthermophilic protein-degrading member of the *Thermoplasmatales*¹⁷, and is adjacent to two copies of gingipain which have the domain architecture of gingipain in *A. boonei*¹⁸ (Fig. 3c and Supplementary Fig. 9). However, 15 of the 16 gingipain copies from all three MBG-D SAGs are monophyletic and distinct from other groups (Supplementary Fig. 9). Cysteine peptidase activity requires chemically reducing, moderate-pH environments with high calcium concentrations¹⁹, which are conditions commonly found in marine sediments. MCG_E09 has two copies of M19, one of the few peptidases known to target D-amino acids. The D enantiomer is highly abundant in the peptidoglycan of bacterial cell walls, which comprise the most persistent sedimentary detrital matter²⁰. MCG_E09 may therefore be specially adapted to degrade these recalcitrant components of cell walls. A comparison with all 4,888 genomes in the IMG database²¹ shows that only mesophilic or moderately thermophilic protein-degrading bacteria share all of

Table 1 | Assembly statistics for each SAG

Statistic	SCGC AB-539-C06	SCGC AB-539-E09	SCGC AB-539-N05	SCGC AB-540-F20
Total generated sequence data (Mb)	491	1,437	2,666	440
Assembly size (Mb)*	0.593	0.627	0.801	1.037
Number of contigs*	104	73	99	172
Maximum contig length (kb)*	48.2	59.3	73.3	77.0
N50 value (kb)*	10.4	27.3	27.1	12.9
Number of predicted genes*	792	787	879	1,272
Detected tRNAs	9	7	22	13
Predicted genome size (tRNA-based)† (Mb)	3.03	4.12	1.67	3.67
Achieved genome coverage (tRNA-based)† (%)	19.6	15.2	48.0	28.3
Detected CSCG	21	18	31	14
Predicted genome size (CSCG-based)‡ (Mb)	1.24	1.53	1.14	3.26
Achieved genome coverage (CSCG-based)‡ (%)	47.8	41.0	70.3	31.8

*Only contigs longer than 1,000 base pairs were considered.

†Based on expected average number of 46 transfer RNAs per genome (Supplementary Fig. 6).

‡Based on 44 determined CSCGs (conserved single-copy genes) common to all archaeal genomes (Supplementary Fig. 6).

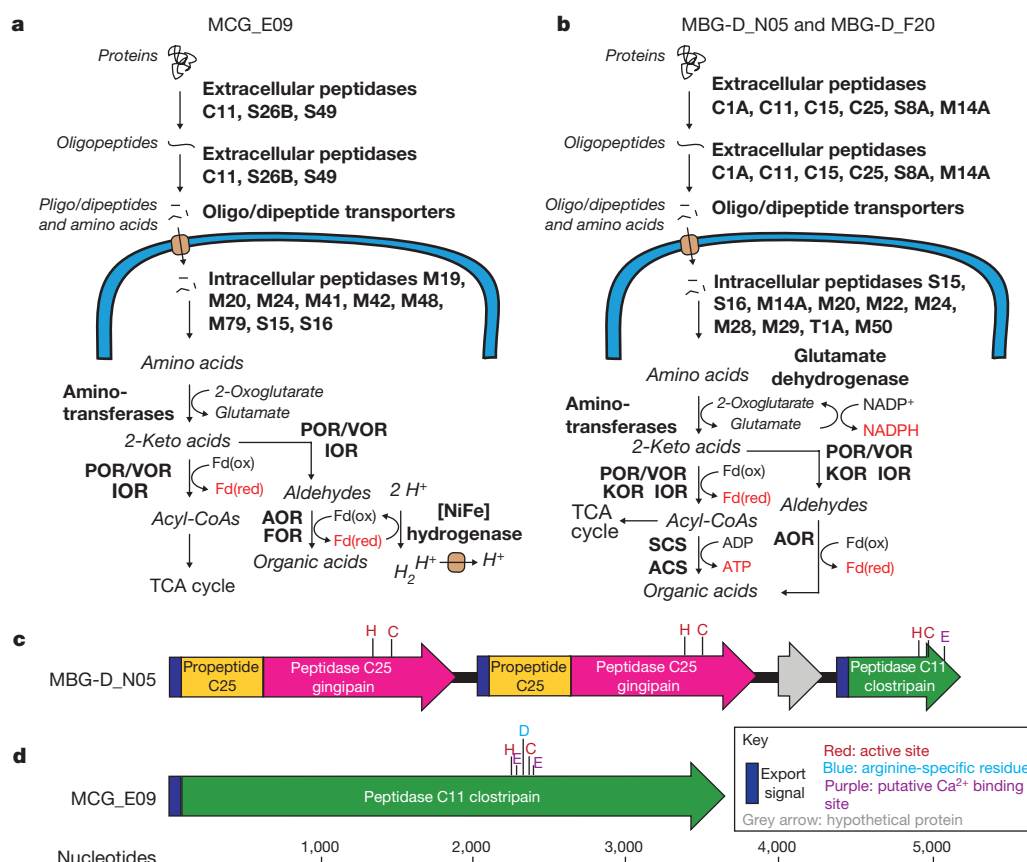


Figure 3 | Proposed protein degradation pathway for MCG_E09 (a) and MBG-D_N05 and MBG-D_F20 (b), and gene architecture for selected extracellular peptidases (c, d). Substrates and products are in black *italic* font, energetic molecules are red, enzymes are in black **bold** font, and blue lines indicate the cell membrane. ACS, acetyl-CoA synthetase; SCS, succinyl-CoA synthetase. Other acronyms are defined in the text. **c, d.** Gene architecture for gingipain and clostripain in MBG-D_N05 (c), and clostripain in MCG_E09 (d). MBG-D_C06 had a partial representation of the pathways present in b.

the MCG and MBG-D peptidases (Supplementary Table 10). Thus, the MCG and MBG-D single cells are probably capable of degrading the detrital proteins that are present in the Aarhus Bay sediment⁸.

All four single cells contained di- and tripeptide transporters as well as genomic pathways for the intracellular breakdown of amino acids. These include aminotransferases, ATP-yielding acetyl-CoA synthetase (in MBG-D), as well as ferredoxin-reducing oxidoreductases specific for aldehydes (AOR, absent in MBG-D_C06), formaldehyde (FOR, only in MCG_E09) and pyruvate/2-ketoisovalerate (POR/VOR, in MCG_E09 and MBG-D_N05), which are intermediates in the breakdown of non-aromatic amino acids in hyperthermophilic archaea²². MCG_E09 and MBG-D_F20 also have indolepyruvate ferredoxin oxidoreductase (IOR), which targets intermediates of aromatic amino acid breakdown²². These oxidoreductases are highly oxygen-labile and use ferredoxin, which has the lowest redox potential of all known electron carriers (-500 mV)²³. Tungsten (the AOR and FOR cofactor) has the lowest redox potential of any biologically relevant metal ligand, is plentiful in the Earth's crust (but not in sea water), and confers slower kinetics than observed for the more common molybdenum-containing oxidoreductases²⁴. Most genomes in the IMG database with Blast hits ($E < 10^{-5}$) to all the ferredoxin-dependent oxidoreductases present in MCG_E09 and MBG-D_N05 (the other two SAGs had fewer oxidoreductases) were hyperthermophilic protein-degrading archaea, with hyperthermophilic protein-degrading bacteria making up most of the rest (Supplementary Table 11). The presence of oxidoreductases normally associated with hyperthermophiles in MCG and MBG-D, which inhabit permanently cold sediments ($2-16^{\circ}\text{C}$ seasonally²⁵), may be ancestral or confer enhanced molecular stability in this reducing, energy-limited environment.

Reduced ferredoxin produced during protein degradation may be used to create a proton motive force in MCG_E09 by its membrane-bound [NiFe]-hydrogenase, analogously to the mechanism present in *Pyrococcus* sp.²⁶ (Supplementary Table 8 and Fig. 3a). MBG-D_N05 and MBG-D_F20 contain heterodisulphide reductase subunits A, B

and C (*hdrABC*), methyl-viologen-reducing hydrogenase subunits A, G and D (*mvhAGD*), and N5-methyltetrahydromethanopterin-coenzyme M methyltransferase subunits A and H (*mtrAH*) (Supplementary Table 8). In some methanogens, these enzymes couple hydrogenotrophic methane production to the creation of a sodium motive force that drives ATP formation²⁷, but the enzymes are also found in non-methanogenic microorganisms. MBG-D_N05 may therefore have a sodium-based energy conservation mechanism.

We observed high extracellular peptidase activity consistent with gingipain and clostripain at 600-cm depth in Aarhus Bay sediments. Gingipain substrates indicated V_{\max} (velocity of enzyme-catalysed reaction at saturating substrate concentrations) = $9.9 \pm 1.6 \mu\text{mol 7-amino-4-methylcoumarin (AMC) h}^{-1} \text{g}^{-1}$ sediment and K_m (Michaelis constant) = $51 \pm 30 \mu\text{M}$ substrate. Clostripain substrates indicated V_{\max} = $15 \pm 5.8 \mu\text{mol AMC h}^{-1} \text{g}^{-1}$ sediment, K_m = $188 \pm 153 \mu\text{M}$ substrate (Supplementary Fig. 10). Leucyl aminopeptidase, which to our knowledge is the only other peptidase substrate that has been assayed in marine sediments²⁸, had much lower potential activity (Supplementary Fig. 10). Archaeal peptidases seem to be numerous in marine sediments, as metagenomes from two geographically disparate marine sediments (California, Gold ID Gm00260, and Alaska, Gold ID Gm00257) are replete with homologues of all extracellular peptidases found in the single cells (up to 2.3 peptidase homologues per genome; $E \leq 10^{-10}$; Supplementary Table 12). This type of extracellular protein degradation within archaea therefore seems to be active, abundant and geographically widespread.

Archaea degrade detrital proteins in extreme environments¹⁷. The partial genomes obtained for MCG and MBG-D suggest that archaea have a similar function in cold anoxic marine sediments, which comprise the largest organic carbon sink on Earth²⁹. Each partial genome contains genes for complete degradation pathways of extracellular proteins, including enzymes whose homologues occur together only in cultured protein-degrading prokaryotes. MCG may represent a new phylum in the Archaea and MBG-D may represent a new order in the

Euryarchaeota because (1) their evolutionarily conserved genes place them basal to established phyla and orders; (2) their environmental distributions differ greatly from their nearest neighbours, Thaumarchaeota, Aigarchaeota and Thermoplasmatales, which are primarily found in oxic and/or hot environments^{4,30}; and (3) MCG and MBG-D single-cell genomes seem to be capable of exogenous protein degradation in cold anoxic environments, a process that has never been observed in other archaea. More single-cell genomes or cultures from these and other uncultured groups may further establish them as taxonomic levels that should be given more formal and accurate names than MCG and MBG-D. The ubiquity and frequent dominance of these archaeal groups, as well as the high abundance and potential activity of the type of peptidases that they encode, indicate the importance of these archaea in protein remineralization in marine sediments. However, the broad 16S rRNA gene diversity within these archaeal groups⁹ indicates that their impacts on marine biogeochemical cycles probably extend beyond their involvement in detrital protein degradation.

METHODS SUMMARY

A sediment core was collected on 22 March 2011, in a shallow gas area at Aarhus Bay, Denmark (56° 9' 35.889 N, 10° 28' 7.893 E), water depth 16.3 m and 2.5 °C (Supplementary Fig. 1). Cells were extracted from 10 cm sediment depth by ultrasonic treatment followed by density gradient centrifugation. Single-cell sorting, whole-genome amplification, and PCR screening of single cells were performed at the Bigelow Laboratory Single Cell Genomics Center (SCGC; <http://www.bigelow.org/scgc>) by FACS using the SYTO-9 DNA stain. The sorted cells were lysed using five cycles of freeze–thaw, followed by further lysis and DNA denaturing with cold KOH. Genomic DNA from the lysed cells was amplified using MDA, resulting in SAGs. MDA products were screened by quantitative PCR with primer sets targeting 16S rRNA genes. Sequencing of the SAGs was performed using the 454-GS-FLX Titanium, the Illumina HiSeq 2000, and the Ion Torrent PGM platforms. Amplicon sequencing (Supplementary Fig. 2) as well as quantitative PCR was performed on DNA from station M1, a sampling site in close proximity to the SAG sampling site (Supplementary Fig. 1). Two different archaea-targeted primer pairs were used for PCR amplification of 16S rRNA gene fragments and subsequent 454 pyrosequencing. The quantitative PCR was performed by using published primer sets for archaea, bacteria and the MCG group as well as newly designed MBG-D primers (Supplementary Table 2). Extracellular peptidase activities were assayed with fluorogenic substrates. L-leucine-7-amino-5-methylcoumarin (Leu-AMC), Z-Phe-Arg-AMC and Z-Phe-Val-Arg-AMC were used as substrates for leucyl aminopeptidase, gingipain and clostripain, respectively. Autoclaved sediment served as a killed control for each substrate and concentration.

Full Methods and any associated references are available in the online version of the paper.

Received 17 October 2012; accepted 20 February 2013.

Published online 27 March 2013.

- Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C. & D'Hondt, S. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl Acad. Sci. USA* **109**, 16213–16216 (2012).
- Schippers, A., Köweler, G., Höft, C. & Teichert, B. M. A. Quantification of microbial communities in forearc sediment basins off Sumatra. *Geomicrobiol. J.* **27**, 170–182 (2010).
- Brochier-Armanet, C., Boussau, B., Gribaldo, S. & Forterre, P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Rev. Microbiol.* **6**, 245–252 (2008).
- Nunoura, T. *et al.* Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* **39**, 3204–3223 (2011).
- Jarrell, K. F. *et al.* Major players on the microbial stage: why archaea are important. *Microbiology* **157**, 919–936 (2011).
- Biddle, J. F. *et al.* Heterotrophic archaea dominate sedimentary subsurface ecosystems off Peru. *Proc. Natl Acad. Sci. USA* **103**, 3846–3851 (2006).
- Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J. & Peterson, M. J. Molecular indicators of diagenetic status in marine organic matter. *Geochim. Cosmochim. Acta* **61**, 5363–5369 (1997).
- Pedersen, A.-G. U., Thomsen, T. R., Lomstein, B. A. & Jørgensen, N. O. G. Bacterial influence on amino acid enantiomerization in a coastal marine sediment. *Limnol. Oceanogr.* **46**, 1358–1369 (2001).
- Kubo, K. *et al.* Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *ISME J.* **6**, 1949–1965 (2012).

- Holmkvist, L., Ferdman, T. G. & Jørgensen, B. B. A cryptic sulfur cycle driven by iron in the methane zone of marine sediment (Aarhus Bay, Denmark). *Geochim. Cosmochim. Acta* **75**, 3581–3599 (2011).
- Raghunathan, A. *et al.* Genomic DNA amplification from a single bacterium. *Appl. Environ. Microbiol.* **71**, 3342–3347 (2005).
- Stepanaukas, R. & Sieracki, M. E. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl Acad. Sci. USA* **104**, 9052–9057 (2007).
- Li, P. *et al.* Genetic structure of three fosmid-fragments encoding 16S rRNA genes of the Miscellaneous Crenarchaeotal Group (MCG): implications for physiology and evolution of marine sedimentary archaea. *Environ. Microbiol.* **14**, 467–479 (2012).
- Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **40**, D343–D350 (2012).
- Barrett, A. J. & Rawlings, N. D. Evolutionary lines of cysteine peptidases. *Biol. Chem.* **382**, 727–733 (2001).
- Labrou, N. E. & Rigden, D. J. The structure–function relationship in the clostripain family of peptidases. *Eur. J. Biochem.* **271**, 983–992 (2004).
- Reysenbach, A. *et al.* A ubiquitous thermophilic archaeon from deep-sea hydrothermal vents. *Nature* **442**, 444–447 (2006).
- Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–D301 (2012).
- Kembhavi, A. A., Buttle, D. J. & Barrett, A. J. Clostripain: characterization of the active site. *FEBS Lett.* **283**, 277–280 (1991).
- Lomstein, B. A., Langerhuus, A. T., Hondt, S. D., Jørgensen, B. B. & Spivack, A. J. Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* **484**, 101–104 (2012).
- Markowitz, V. M. *et al.* IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res.* **40**, D123–D129 (2012).
- Schut, G. J., Menon, A. L. & Adams, M. W. W. 2-keto acid oxidoreductases from *Pyrococcus furiosus* and *Thermococcus litoralis*. *Methods Enzymol.* **331**, 144–158 (2001).
- Hall, D. O., Cammack, R. & Rao, K. K. Role for ferredoxins in the origin of life and biological evolution. *Nature* **233**, 136–138 (1971).
- Johnson, M. K., Rees, D. C. & Adams, M. W. W. Tungstoenzymes. *Chem. Rev.* **96**, 2817–2840 (1996).
- Dale, A. W. *et al.* Seasonal dynamics of the depth and rate of anaerobic oxidation of methane in Aarhus Bay (Denmark) sediments. *J. Mar. Res.* **66**, 127–155 (2008).
- Sapra, R., Bagramyan, K. & Adams, M. A simple energy-conserving system: Proton reduction coupled to proton translocation. *Proc. Natl Acad. Sci. USA* **100**, 7545–7550 (2003).
- Thauer, R., Kaster, A.-K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Rev. Microbiol.* **6**, 579–591 (2008).
- Coolen, M. J. L. & Overmann, J. Functional exoenzymes as indicators of metabolically active bacteria in 124,000-year-old sapropel layers of the Eastern Mediterranean Sea. *Appl. Environ. Microbiol.* **66**, 2589–2598 (2000).
- Hedges, J. I. & Keil, R. G. Sedimentary organic matter preservation: an assessment and speculative synthesis. *Mar. Chem.* **49**, 81–115 (1995).
- Brochier-Armanet, C., Gribaldo, S. & Forterre, P. Spotlight on the Thaumarchaeota. *ISME J.* **6**, 227–230 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank the captain and crew of the R/V *Tyra* for sampling; T. B. Sogaard, A. Stentebjerg and B. Poulsen for technical work; F. Löffler for laboratory space; and D. Kirchman and S. Hallam for sharing their unpublished metagenomic data sets. This work was funded by the Danish National Research Foundation, the German Max Planck Society, NSF Center for Dark Energy Biosphere Investigations NSF-157595 (K.G.L.), The Danish Council for Independent Research–Natural Sciences (D.G.P.), the Villum Kann Rasmussen Foundation, an EU Marie Curie fellowship (M.A.L.), the German Research Foundation (S.L.) and the USA National Science Foundation awards EF-826924, OCE-821374 and OCE-1019242 (R.S.).

Author Contributions K.G.L., L.S., D.G.P., K.U.K., R.S., A.S. and B.B.J. worked together to design experiment and develop the method for single cell sorting from sediments. K.G.L. wrote the main paper and developed the protein degradation hypothesis. L.S. wrote the Supplementary Information, designed and performed bioinformatic analyses. K.G.L. and L.S. performed phylogenetic tests. K.G.L. and D.G.P. reconstructed metabolic pathways with SAG genes. R.S. performed cell sorting and amplification. S.K., S.L., D.G.P. and L.S. developed protocols for cell separation from sediments. K.U.K. performed and analysed 16S rRNA gene amplicon sequencing; M.A.L., L.S. and K.G.L. performed quantitative PCR; A.D.S. performed enzyme activity measurements; and M.R. gave bioinformatic support and added quality control tests. A.S. and B.B.J. obtained the major funding for this work. All co-authors commented on and provided substantial edits to the manuscript.

Author Information This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank as Thaumarchaeota archaeon SCGC AB-539-E09 (accession number ALXK00000000), Thermoplasmatales archaeon SCGC AB-539-C06 (AOSH00000000), Thermoplasmatales archaeon SCGC AB-539-N05 (ALXL00000000) and Thermoplasmatales archaeon SCGC AB-540-F20 (AOSI00000000). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.G.L. (klloyd@utk.edu).

METHODS

Sequencing of SAGs, sequence assembly and annotation. A sediment core was collected on 22 March 2011, in a shallow gas area at Aarhus Bay, Denmark (56° 9' 35.889 N, 10° 28' 7.893 E), water depth 16.3 m and 2.5 °C (Supplementary Fig. 1). Cells were extracted from 10 cm sediment depth by ultrasonic treatment followed by density gradient centrifugation. Single-cell sorting, whole-genome amplification, and PCR screening of single cells were performed at the Bigelow Laboratory Single Cell Genomics Center (SCGC; <http://www.bigelow.org/scgc>) by FACS using the SYTO-9 DNA stain. The sorted cells were lysed using five cycles of freeze–thaw, followed by further lysis and DNA denaturing with cold KOH. Genomic DNA from the lysed cells was amplified using MDA, resulting in SAGs. MDA products were screened by quantitative PCR with primer sets targeting 16S rRNA genes. To obtain sufficient quantity of DNA for sequencing, some SAGs were re-amplified by a second round of MDA at the SCGC. Sequencing with the 454-GS-FLX Titanium and the Illumina HiSeq 2000 was performed at GATC Biotech (Germany) with re-amplified SAGs as template. Sequencing with the Ion Torrent PGM platform was performed at Aarhus University using the original (non-re-amplified) SAGs as template. The sequence data for the individual SAGs are summarized in Supplementary Table 4.

Reads from the 454-GS-FLX Titanium and the Ion Torrent PGM platforms were quality trimmed using the prinseq-lite.pl script³¹. The 454 and Ion Torrent reads were de-replicated using cd-hit-454³² and subsequently assembled using gsAssembler version 2.6 (Roche). In parallel, we used the SPAdes assembler version 2.2.1³³ to assemble the Illumina reads of SCGC AB-539-E09 and SCGC AB-539-N05, as well as the Ion Torrent reads of SCGC AB-539-C06 and SCGC AB-540-F20. The gsAssembler and SPAdes assemblies of each SAG were combined using Sequencher version 5.0.1 (Genecodes). The SAG assemblies were auto-annotated using the Joint Genome Institute (JGI) IMG-ER pipeline²¹. Annotations were manually curated using GenDB³⁴ supplemented by JCoast³⁵. All peptides were aligned in ARB³⁶ against the Merops alignment¹⁴ (Supplementary Fig. 9).

Estimation of genome size and purity control of SAGs. The genome sizes of the SAGs were estimated using a conserved single-copy gene (CSCG) analysis similar to ref. 37. CSCGs present in all archaea were identified using the JGI IMG site³⁸. The CSCG-based approach was supplemented by a tRNA-based approach³⁹ where SAG tRNA numbers were compared to the numbers of tRNAs of complete archaeal genomes.

The purity of the SAGs was tested by PCR-screening for archaeal and bacterial 16S rRNA genes using multiple primer sets. Additionally, all predicted ORFs were checked with Blastp⁴⁰ against NCBI-nr for amino acid identity observations of over 96% to detect common contaminants. Finally, contigs showing a k-mer pattern divergent from the rest of the SAG sequences⁴¹ were manually checked for a possible contamination by examining their closest Blastp hits.

Phylogenetic reconstruction using conserved single-copy genes and 16S rRNA genes. Similar to the studies performed by refs 42 and 43, we used archaeal CSCGs for inferring the phylogenetic affiliation of the inspected SAGs. The identified CSCG amino acid sequences of completed archaeal genomes and our SAGs were extracted from the JGI IMG-ER site. The sequences were individually aligned using MAFFT version 6.864⁴⁴. All alignments were manually curated using ARB³⁶. The alignments were concatenated using the Perl script catfasta2phym.pl (<http://www.abc.se/~nylander/catfasta2phym.pl>). Maximum likelihood bootstrap trees were calculated using RAxML-HP2 (RAxML version 7.2.7⁴⁵) as provided by the CIPRES cluster (<http://www.phylo.org/46>). The 16S rRNA gene tree was created using RAxML-HP2⁴⁵ at the CIPRES cluster⁴⁶ (Supplementary Fig. 8).

Aarhus Bay 16S rRNA amplicon sequencing and quantitative PCR. A sediment core was taken at station M1, (56° 07.07' N, 10° 20.80' E; see Supplementary Fig. 1), a well-characterized site near the site from which SAGs were derived in Aarhus Bay⁴⁷. DNA was extracted⁴⁸ from five depths. Two different archaea-targeted primer pairs^{49–52} were used for PCR amplification of 16S rRNA gene fragments. The resultant PCR products were sequenced on a 454-GS-FLX Titanium machine as previously described⁵³. Sequence analysis was performed in Mothur⁵⁴. Sequence reads were classified according to Silva taxonomy (release 102 (ref. 55)) and new MCG subgroups⁹ (Supplementary Fig. 2).

The DNA for quantitative PCR was extracted from the same sediment used for 16S rRNA amplicon sequencing, station M1, using a novel, chemical lysis-based method (M.A.L., manuscript in preparation). The quantitative PCR was performed with primers listed in Supplementary Table 2 (Supplementary Fig. 2). Primers for MBG-D (Supplementary Tables 2 and 3) were designed using PRIMROSE 2.17 (ref. 56) and 839 MBG-D sequences included in the SILVA SSURF database release 106 (ref. 55).

Enzymatic activity assays. Extracellular peptidase activities were assayed using fluorogenic substrates according to a protocol loosely based on that of ref. 57 (Supplementary Fig. 10). L-leucine-7-amino-4-methylcoumarin (Leu-AMC, Sigma-Aldrich) was used to assay leucyl aminopeptidase activity, Z-Phe-Arg-AMC

(Sigma-Aldrich) was used to assay gingipain⁵⁸ and Z-Phe-Val-Arg-AMC (Bachem) was used to assay clostripain⁵⁹. Sediments were collected by gravity core from Aarhus Bay site M1. Sediments from 600 to 630 cm below sea floor were homogenized, placed in 5-ml serum vials (~0.5 g each), mixed with 4.0 ml anoxic artificial sea water (salinity 30 practical salinity units, pH 7.8) and the precise mass of wet sediment was recorded. Serum vials were then capped, purged with N₂ and vortexed to mix completely. For each substrate and concentration, autoclaved sediment was used as a killed control. Immediately after vortexing, 2 ml slurry was removed to measure initial fluorescence.

To measure fluorescence, each sediment subsample was transferred into a microcentrifuge vial and centrifuged for 10 min at 9,300g. One millilitre of supernatant was removed and transferred to a methacrylate 1.5-ml fluorescence cuvette. Fluorescence was measured with a Promega Quantifluor ST fluorimeter. To calibrate fluorescence values, 7-amino-4-methylcoumarin (AMC) was added directly to sediment slurries, mixed thoroughly, and then processed in the same way as samples for enzyme assays. Enzyme activities are reported as rates of AMC liberation per hour per gram wet sediment.

Sediments were incubated in the dark in a shaking incubator for approximately 8 h (precise time was recorded). After incubation, fluorescence was measured as described above. Enzyme activity was calculated from the change in fluorescence for each sample divided by the incubation time. Kinetic parameters were calculated using the R statistical package⁶⁰ by a nonlinear least-squares fit to the activity data (Supplementary Fig. 10).

- Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
- Niu, B., Fu, L., Sun, S. & Li, W. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* **11**, 187 (2010).
- Bankevich, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
- Meyer, F. et al. GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* **31**, 2187–2195 (2003).
- Richter, M. et al. JCoast—A biologist-centric software tool for data mining and comparison of prokaryotic (meta)genomes. *BMC Bioinformatics* **9**, 177 (2008).
- Ludwig, W. et al. ARB: a software environment for sequence data. *Database* **32**, 1363–1371 (2004).
- Woyke, T. et al. Assembling the marine metagenome, one cell at a time. *PLoS ONE* **4**, e5299 (2009).
- Markowitz, V. M. et al. The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* **34**, D344–D348 (2006).
- Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A Program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 0955–0964 (1997).
- Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Swan, B. K. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**, 1296–1300 (2011).
- Matte-Tailliez, O., Brochier, C., Forterre, P. & Philippe, H. Archaeal phylogeny based on ribosomal proteins. *Mol. Biol. Evol.* **19**, 631–639 (2002).
- Brochier, C., Forterre, P. & Gribaldo, S. An emerging phylogenetic core of Archaea: phylogenies of transcription and translation machineries converge following addition of new genome sequences. *BMC Evol. Biol.* **5**, 36 (2005).
- Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
- Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**, 758–771 (2008).
- Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop 1–8 (2010).
- Jørgensen, B. B. Case study—Århus Bay. *Eutrophication in Coastal Marine Systems* 137–154 (American Geophysical Union, 1996).
- Kjeldsen, K. U. et al. Diversity of sulfate-reducing bacteria from an extreme hypersaline sediment, Great Salt Lake (Utah). *FEMS Microbiol. Ecol.* **60**, 287–298 (2007).
- DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
- Stahl, D. A. & Amann, R. *Development and Application of Nucleic Acid Probes* (Wiley, 1991).
- Takai, K. E. N. & Horikoshi, K. Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Appl. Environ. Microbiol.* **66**, 5066–5072 (2000).
- Teske, A. & Sørensen, K. B. Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J.* **2**, 3–18 (2008).
- Larsen, N. et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE* **5**, e9085 (2010).
- Schloss, P. D. et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Society* **75**, 7537–7541 (2009).
- Pruesse, E. et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).

56. Ashelford, K. E., Weightman, A. J. & Fry, J. C. PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Res.* **30**, 3481–3489 (2002).
57. King, G. M. Characterization of β -glucosidase activity in intertidal marine sediments. *Appl. Environ. Microbiol.* **51**, 373–380 (1986).
58. Nakayama, K., Kadowaki, T., Okamoto, K. & Yamamoto, K. Construction and characterization of arginine-specific cysteine proteinase (Arg-gingipain)-deficient mutants of *Porphyromonas gingivalis*: Evidence for significant contribution of Arg-gingipain to virulence. *J. Biol. Chem.* **270**, 23619–23626 (1995).
59. Rauber, P., Walker, B., Stone, S. & Shaw, E. Synthesis of lysine-containing sulphonium salts and their properties as proteinase inhibitors. *Biochem. J.* **250**, 871 (1988).
60. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing <http://www.rproject.org> (2012).

Diverging neural pathways assemble a behavioural state from separable features in anxiety

Sung-Yon Kim^{1,2*}, Avishek Adhikari^{1*}, Soo Yeun Lee^{1,3}, James H. Marshel¹, Christina K. Kim^{1,2}, Caitlin S. Mallory^{1,2}, Maisie Lo¹, Sally Pak¹, Joanna Mattis^{1,2}, Byung Kook Lim⁴, Robert C. Malenka⁴, Melissa R. Warden¹, Rachael Neve⁵, Kay M. Tye^{1,5} & Karl Deisseroth^{1,2,3,4,6}

Behavioural states in mammals, such as the anxious state, are characterized by several features that are coordinately regulated by diverse nervous system outputs, ranging from behavioural choice patterns to changes in physiology (in anxiety, exemplified respectively by risk-avoidance and respiratory rate alterations)^{1,2}. Here we investigate if and how defined neural projections arising from a single coordinating brain region in mice could mediate diverse features of anxiety. Integrating behavioural assays, *in vivo* and *in vitro* electrophysiology, respiratory physiology and optogenetics, we identify a surprising new role for the bed nucleus of the stria terminalis (BNST) in the coordinated modulation of diverse anxiety features. First, two BNST subregions were unexpectedly found to exert opposite effects on the anxious state: oval BNST activity promoted several independent anxious state features, whereas anterodorsal BNST-associated activity exerted anxiolytic influence for the same features. Notably, we found that three distinct anterodorsal BNST efferent projections—to the lateral hypothalamus, parabrachial nucleus and ventral tegmental area—each implemented an independent feature of anxiolysis: reduced risk-avoidance, reduced respiratory rate, and increased positive valence, respectively. Furthermore, selective inhibition of corresponding circuit elements in freely moving mice showed opposing behavioural effects compared with excitation, and *in vivo* recordings during free behaviour showed native spiking patterns in anterodorsal BNST neurons that differentiated safe and anxiogenic environments. These results demonstrate that distinct BNST subregions exert opposite effects in modulating anxiety, establish separable anxiolytic roles for different anterodorsal BNST projections, and illustrate circuit mechanisms underlying selection of features for the assembly of the anxious state.

Animals encounter environmental conditions that require rapid switching among different behavioural states to increase the likelihood of survival and reproduction. Such states consist of a constellation of changes coordinated by distinct modalities of nervous system output^{1,2}, and understanding this behavioural-state assembly from diverse features is of fundamental interest. A well-studied example is fear, in which the amygdala is thought to modulate various aspects of fear expression by multiple downstream targets^{1–4}. Here we tested whether specific diverging projections causally recruit separable features to coordinate a behavioural state, focusing on anxiety as a state not only important in normal and pathological behaviour⁵, but also exhibiting many disparate features that are quantifiable in mice.

Evidence from anatomical^{6–8}, behavioural^{9,10} and neuroimaging studies^{11,12} has implicated the BNST in pathological and adaptive anxiety; for example, lesions of the dorsal BNST, henceforth referred to as BNST, have been reported to decrease anxiety-like behaviour^{9,10}. To test this finding further, we infused glutamate receptor antagonists

into the BNST before the elevated-plus maze (EPM)¹⁰ test in mice (Fig. 1a; histology in Supplementary Figs 1–3). This intervention increased open-arm exploration ($P < 0.01$, see Supplementary Information for statistical analysis; Fig. 1a) without altering locomotion (Supplementary Fig. 4; such increased exploration of open spaces, to which mice exhibit innate aversion, is thought to represent reduced anxiety-like behaviour¹³). We next optogenetically inhibited the BNST using an enhanced form of the inhibitory *Natronomonas pharaonis* halorhodopsin (eNpHR3.0)¹⁴ and delivery of yellow light to the BNST (eNpHR3.0:BNST somata; Fig. 1b); increased exploration of open spaces in the EPM test and open field test (OFT) was observed (Fig. 1b and Supplementary Fig. 5a, b), indicating anxiolysis. Conversely, stimulation of BNST somata with the excitatory channelrhodopsin-2 (ChR2) increased behavioural measures of anxiety in both assays (ChR2:BNST somata; Supplementary Fig. 6). To test whether this manipulation modulated physiological manifestations of anxiety, we stimulated BNST somata while monitoring respiratory rate; hyperventilation is linked to increased anxiety in humans^{15,16} and rodents (Supplementary Fig. 7), and the BNST is known to project to respiratory centres^{17,18}. Indeed, increased respiratory rate was observed (Supplementary Fig. 6d). Together these results suggest that activity in the BNST drives an anxiety-like state, consistent with most previous studies⁹.

However, these results may not provide a complete picture of the BNST, which contains several subregions^{7,19,20}. We next targeted the oval nucleus of the BNST (hereafter termed ovBNST), by introducing a Cre-dependent eNpHR3.0 virus into the BNST of dopamine receptor 1a (Drd1a::Cre) mice that show restricted Cre expression in the ovBNST (eNpHR3.0:ovBNST; Fig. 1c). Yellow light delivery in eNpHR3.0:ovBNST mice decreased avoidance of EPM open arms ($P < 0.0001$; Fig. 1d) and the OFT centre ($P < 0.001$; Supplementary Fig. 5c). The same manipulation also decreased respiratory rate ($P < 0.05$; Fig. 1e). Conversely, stimulating the ovBNST with ChR2 increased both behavioural and physiological measures of anxiety (ChR2:ovBNST; Supplementary Fig. 8). These results suggested an anxiogenic role for the ovBNST, and were consistent with the results obtained by modulating the entire BNST (Fig. 1a, b).

We next investigated the function of basolateral amygdala (BLA) inputs to the BNST, as the BLA is a region implicated in anxiety^{2,9} that projects to the BNST⁷. Mice expressing eNpHR3.0–eYFP in BLA pyramidal neurons displayed enhanced yellow fluorescent protein (eYFP)⁺ fibres projecting to the region of the BNST surrounding the ovBNST, hereafter referred to as anterodorsal BNST, or adBNST⁷ (eNpHR3.0:BLA-adBNST; Fig. 1f). Surprisingly, inhibiting the BLA-adBNST projection increased avoidance of EPM open arms ($P < 0.01$; Fig. 1g) and the OFT centre ($P < 0.01$; Supplementary Fig. 5e), and also increased respiratory rate ($P < 0.01$; Fig. 1h). Conversely, stimulating BLA

¹Department of Bioengineering, Stanford University, Stanford, California 94305, USA. ²Neurosciences Program, Stanford University, Stanford, California 94305, USA. ³CNC Program, Stanford University, Stanford, California 94305, USA. ⁴Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA. ⁵Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁶Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA.

*These authors contributed equally to this work.

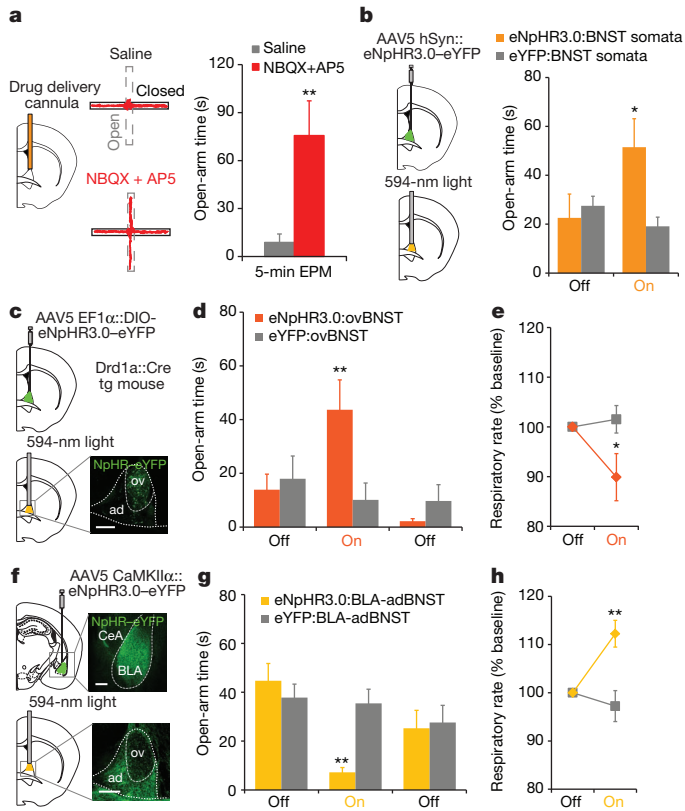


Figure 1 | Functional heterogeneity within the dorsal BNST. **a**, Cannula for drug infusion; NBQX (2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f]quinoxaline-7-sulphonamide) plus AP5 (D(-)-2-amino-5-phosphonopentanoic acid) increased open-arm time in the EPM test ($n = 5$ for each). **b**, eNpHR3.0:BNST somata mice were bilaterally implanted with fibre-optics above the BNST. Light increased open-arm time in the EPM test ($n = 8$ eNpHR3.0, $n = 8$ eYFP). eNpHR3.0-eYFP was under the control of the human synapsin 1 promoter (adeno-associated virus type 5 (AAV5) human synapsin promoter fragment (hSyn)::eNpHR3.0-eYFP). **c**, eNpHR3.0:ovBNST mice received bilateral light. ovBNST-restricted expression was obtained with Cre-dependent eNpHR3.0 adeno-associated virus in Drd1a::Cre transgenic (tg) mice. eNpHR3.0-eYFP was under the control of the EF1α promoter. **d**, **e**, Light delivery to ovBNST of eNpHR3.0:ovBNST mice increased open-arm time in the EPM test ($n = 7$ eNpHR3.0, $n = 8$ eYFP) (**d**) and decreased respiratory rate ($n = 7$ eNpHR3.0, $n = 8$ eYFP) (**e**). **f**, eNpHR3.0:BLA-adBNST mice expressing eNpHR3.0 in BLA received bilateral illumination of BLA fibres in adBNST. eNpHR3.0-eYFP was under the control of the CaMKIIα promoter. **g**, **h**, Light in eNpHR3.0:BLA-adBNST mice reduced open-arm time ($n = 11$ eNpHR3.0, $n = 15$ eYFP) (**g**) and increased respiratory rate ($n = 8$ eNpHR3.0, $n = 8$ eYFP) (**h**). Scale bars, 200 μm. Data are mean ± s.e.m. * $P < 0.05$; ** $P < 0.01$. Statistics in Supplementary Information; see also Supplementary Fig. 5.

inputs with Chr2 (Chr2:BLA-adBNST; Fig. 2a) decreased behavioural anxiety measures (Fig. 2b and Supplementary Figs 9a, b and 10) and respiratory rate ($P < 0.05$; Fig. 2c). Because the BLA projection is thought to be excitatory, as confirmed later, these data suggest that adBNST recruitment is anxiolytic, in contrast to the anxiogenic nature of ovBNST activity. Importantly, these effects were not attributable to excitation of BLA fibres in the anterior commissure (Supplementary Fig. 11). As an additional test, considering that a clinically relevant^{21,22} feature of anxiolysis can be positive subjective valence, we asked whether stimulating BLA-adBNST projections could elicit positive conditioning valence (using the real-time place preference (RTPP) task; see Methods), but we did not observe elicited place preference (Fig. 2d).

Having found that driving adBNST afferents decreases avoidance of open spaces and respiratory rate, we next investigated which adBNST outputs might mediate these distinct effects. The adBNST projection to the lateral hypothalamus (LH) was a candidate for mediating

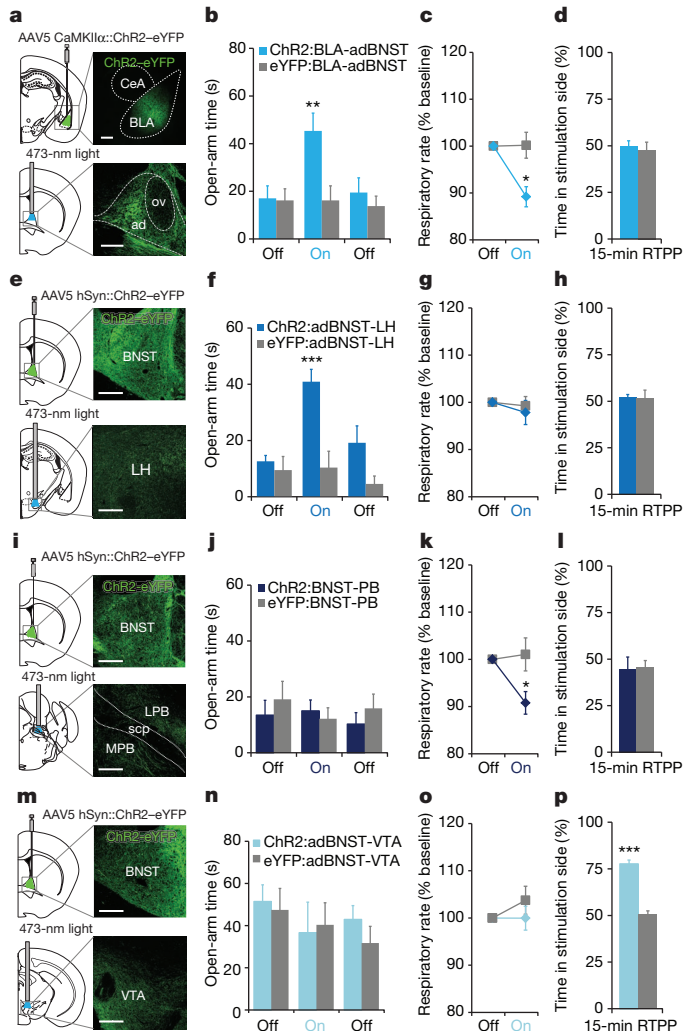


Figure 2 | Distinct adBNST outputs modulate different features related to anxiolysis. **a**, Chr2:BLA-adBNST mice were transduced in BLA, and unilateral fibre-optics implanted above BLA fibres in adBNST. CeA, central amygdala. **b–d**, Light to adBNST increased open-arm time in the EPM test ($n = 11$ Chr2, $n = 12$ eYFP) (**b**) and decreased respiratory rate ($n = 7$ Chr2, $n = 8$ eYFP) (**c**), but did not elicit place preference ($n = 8$ Chr2, $n = 6$ eYFP) (**d**). **e**, Chr2:adBNST-LH mice were transduced in BNST, and unilateral fibre-optics implanted above the lateral hypothalamus. **f–h**, In Chr2:adBNST-LH mice, light increased the open-arm time in the EPM test ($n = 11$ Chr2, $n = 8$ eYFP) (**f**), but did not affect respiratory rate ($n = 9$ Chr2, $n = 10$ eYFP) (**g**) or place preference ($n = 7$ Chr2, $n = 7$ eYFP) (**h**). **i**, Chr2:BNST-PB mice were transduced in BNST, and unilateral fibre-optics implanted in the parabrachial nucleus. LPB, lateral parabrachial nucleus; MPB, medial parabrachial nucleus; scp, superior cerebellar peduncle. **j–l**, Light in Chr2:BNST-PB mice did not influence performance in the EPM test ($n = 7$ Chr2, $n = 7$ eYFP) (**j**), but decreased respiratory rate ($n = 8$ Chr2, $n = 7$ eYFP) (**k**); no effect was seen on place preference ($n = 7$ Chr2, $n = 5$ eYFP) (**l**). **m**, Chr2:adBNST-VTA mice were transduced in the BNST, and unilateral fibre-optics implanted directly above the VTA. **n–p**, Light did not affect the EPM test ($n = 7$ Chr2, $n = 7$ eYFP) (**n**) or respiratory rate ($n = 8$ Chr2, $n = 7$ eYFP) (**o**), but induced robust place preference ($n = 8$ Chr2, $n = 7$ eYFP) (**p**). Scale bars, 200 μm. Data are mean ± s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Statistics in Supplementary Information; see also Supplementary Fig. 9.

decreases in behavioural expression of anxiety, as the lateral hypothalamus receives projections from the adBNST but not the ovBNST^{18,23,24} (Supplementary Fig. 12a), and is required for normal behaviour in the EPM test²⁵. In agreement with this hypothesis, we found that adBNST neurons projecting to the lateral hypothalamus receive BLA input (Supplementary Fig. 12b–d), and that stimulating the adBNST-LH projection decreased avoidance of open spaces in both the EPM test

($P < 0.01$; Fig. 2f) and the OFT ($P < 0.05$; Supplementary Fig. 9c). However, no effects were seen on respiratory rate (Fig. 2g) or the RTPP test (Fig. 2h), suggesting that the adBNST-LH pathway selectively modulates behavioural, but not physiological or appetitive, features of anxiolysis.

We proposed that the adBNST output to the parabrachial nucleus (PB) could mediate the decrease in respiratory rate seen in ChR2:BLA-adBNST mice (Fig. 2c), as the parabrachial nucleus modulates respiration^{2,17,26}. Indeed, in ChR2:BNST-PB mice (Fig. 2i), blue light decreased respiratory rate ($P < 0.05$; Fig. 2k). Furthermore, stimulating the BNST-PB projection attenuated respiratory rate increases in an anxiogenic environment (Supplementary Fig. 13), but did not change behaviour in the EPM or the RTPP tests (Fig. 2j, l). Although both the adBNST and the ovBNST project to the parabrachial nucleus^{18,23,24}, the decreased respiratory rate in ChR2:BNST-PB mice was probably driven by adBNST-PB fibres, as ovBNST activity increased the respiratory rate (Fig. 1e and Supplementary Fig. 8). Finally, we tested the adBNST output to the ventral tegmental area (VTA)^{18,23,24,27,28}. Remarkably, ChR2:adBNST-VTA mice (Fig. 2m) exhibited place preference in the stimulated chamber ($P < 0.001$; Fig. 2p), without affecting anxiety-related risk-avoidance (Fig. 2n) or respiratory rate (Fig. 2o). These data showing complementary roles of different adBNST projections support a model in which populations of adBNST neurons project to distinct downstream structures (lateral hypothalamus, parabrachial nucleus and VTA; Supplementary Fig. 14), modulating different features related to anxiolysis.

We next investigated the intrinsic microcircuitry of the adBNST. To examine connectivity between the BLA and the adBNST, mice expressing ChR2 in the BLA were implanted with a microdrive containing stereotrodes surrounding a fibre-optic in the adBNST (Fig. 3a and Supplementary Fig. 15), allowing simultaneous excitation and recording in awake animals. As expected, excitation of the glutamatergic BLA terminals increased spiking of adBNST single units (Fig. 3b, c), and corresponding whole-cell patch recordings from acute slices revealed that 84% of the adBNST neurons exhibiting both evoked excitatory postsynaptic currents (EPSCs) and inhibitory postsynaptic currents (IPSCs) in voltage clamp (Methods and Supplementary Fig. 16) displayed net excitation in response to BLA input stimulation in current clamp (Fig. 3d–f). Thus, *in vivo* and *in vitro* electrophysiology were concordant in showing that stimulating the BLA-adBNST projection increases adBNST activity, which may be enhanced by local adBNST recurrent excitation (Supplementary Figs 16 and 17). We also characterized local inputs to the adBNST, by recording from adBNST neurons while optically stimulating ovBNST inputs (Fig. 3g). Interestingly, 79% of neurons displayed net inhibition (Fig. 3h, i), consistent with the fact that ovBNST neurons are mostly GABAergic²⁹; by contrast, retrograde tracing experiments showed that the adBNST only weakly projects^{18,23} to the ovBNST (Supplementary Fig. 18). Together, these data support the conclusion that the ovBNST and adBNST exhibit opposing roles in modulating anxiety.

Next, we asked whether the native firing rates of adBNST neurons in freely moving mice encoded aspects of environmental safety, by recording activity with stereotrode arrays in the adBNST during exploration (Fig. 4a, b). Indeed, greater adBNST multiunit activity was observed in safer locations in two models (closed arms of the EPM test and dark compartment of the light–dark test box; Supplementary Fig. 19). To quantify the extent to which adBNST single units differentiated between closed and open arms in the EPM test, we defined an EPM score (see Supplementary Methods and Supplementary Fig. 20), in which a positive score indicates that firing rates are similar between arms of the same type (such as a pair of closed arms), but different across open and closed arms³⁰ (for example, Fig. 4c). This metric allowed calculation of specific EPM performance-related activity for each single unit both in light-on and light-off epochs. Without illumination, a subset of adBNST single units fired preferentially in the closed arms of the EPM test, whereas other units did not exhibit

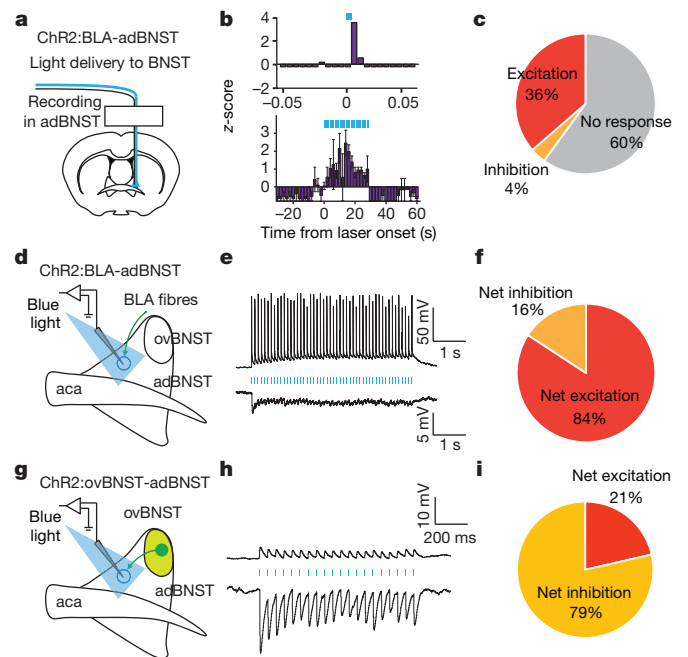


Figure 3 | *In vivo* and *in vitro* electrophysiological assessment of adBNST afferents. **a–f**, Assessment of BLA afferents to adBNST. **a**, ChR2:BLA-adBNST mice were implanted with a microdrive containing eight stereotrodes and a fibre-optic in adBNST to allow simultaneous optogenetic stimulation and recording of adBNST neurons. **b**, Representative peristimulus time histograms of adBNST single units in behaving mice, showing typical responses to 5-ms light pulse (top), and to a 10-Hz light-pulse train for 20 s (bottom). **c**, Excitation was most commonly observed ($n = 55$). **d**, ChR2 was expressed in the BLA; acute slices were prepared from the BNST, and BNST neurons were recorded in current-clamp while optically stimulating BLA afferents. **e**, Representative traces from adBNST neurons ($V_m \approx -60$ mV), displaying excitatory (top) and inhibitory (bottom) responses. **f**, Among adBNST neurons that showed both EPSCs and IPSCs, most were excited at resting potential ($n = 16$ out of 19 neurons; see Supplementary Fig. 16 for voltage-clamp). **g–i**, Electrophysiologically assessed functional connectivity from ovBNST to adBNST (Supplementary Fig. 18 illustrates minimal connectivity in the reverse direction). **g**, ChR2 was expressed in ovBNST using transgenic *Drd1a::Cre* mice; adBNST neurons were recorded while stimulating ovBNST fibres. **h**, Representative current-clamp traces from adBNST neurons ($V_m \approx -60$ mV), exhibiting excitatory (top) and inhibitory (bottom) responses. **i**, Among adBNST neurons that showed both EPSCs and IPSCs, most were inhibited at resting potential ($n = 11$ out of 14 neurons). Data are mean \pm s.e.m. Statistics in Supplementary Information.

preference (Fig. 4c). In fact, every adBNST single unit with a positive EPM score (66% of units) had higher firing rates in the closed arms than in the open arms, whereas simulations predict that if there were no dependence on environmental condition, only 33% of cells would have a positive EPM score, and those would be evenly divided between closed- and open-arm-preferring units (Supplementary Methods).

We then implanted stereotrodes and a fibre-optic in the adBNST of eNpHR3.0:BLA-adBNST mice (Fig. 4a), allowing simultaneous recording and yellow light delivery to the adBNST. Illumination in these mice reduced multiunit activity in the adBNST (Fig. 4d and Supplementary Fig. 21). Finally, we recorded from adBNST single units in eNpHR3.0:BLA-adBNST mice during the EPM test for 20 min, with alternating 1-min light-off and light-on epochs (Fig. 4e), to allow calculation of EPM scores for each single unit in the presence or absence of inhibition of BLA afferents. Suggesting that representation of anxiety-related features in the adBNST may depend on BLA input, we observed that optogenetic inhibition of the BLA-adBNST projection decreased single-unit EPM scores ($P < 0.01$; Fig. 4f, g), and the decrease in EPM scores was higher in cells that had decreases in the firing rate during the illuminated epochs (Supplementary Fig. 22).

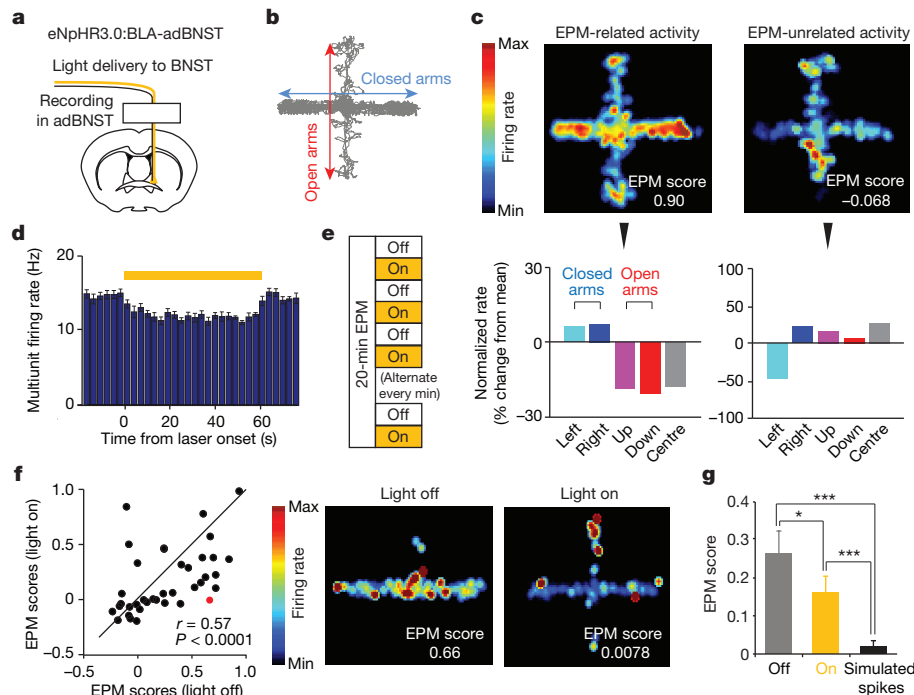


Figure 4 | BNST neurons rely in part on BLA inputs to distinguish safe and anxiogenic locations. **a**, Schematic of *in vivo* recording configuration. **b**, Representative behavioural track tracing from the EPM test. For all EPM figures, horizontal and vertical arms represent closed and open arms, respectively. **c**, Top, spatial firing rate maps of two representative adBNST single units. One unit showed higher activity in closed arms (left), whereas the other did not exhibit preference (right); average normalized firing rates are colour-coded for each pixel of spatial location. Bottom, normalized rates (percentage change from mean firing rate) for each arm for example units. These rates were used to calculate EPM scores (Supplementary Methods and Supplementary Fig. 20); higher EPM scores indicate greater differentiation of closed and open arms. **d**, Light to inhibit the BLA-adBNST projection modestly

These data indicate that native anxiety-related encoding of the EPM environment in the adBNST depends in part on BLA inputs; note that this same manipulation (inhibiting the BLA-adBNST projection) increased anxiety-like behaviour in the EPM test (Fig. 1g), in a manner consistent with causing increased overall anxiety that could deter transitions to the open arm.

Here, we have mapped the role of BNST circuit elements in the assembly and modulation of the anxious behavioural state. We have demonstrated that the ovBNST and adBNST increase and decrease anxiety-related behaviour, respectively; the ovBNST could promote anxiety by suppressing the adBNST (see Supplementary Fig. 23 for summary diagram) or via direct projections to structures such as the central amygdala. We next found that distinct adBNST projections modulate different features of the behavioural state associated with anxiolysis—decreased risk-avoidance behaviour, decreased respiratory rate and positive conditioning valence—which are mediated by projections from the adBNST to the lateral hypothalamus, parabrachial nucleus and VTA, respectively. This arrangement may facilitate modular adaptation of the state itself over development and experience; in principle, by tuning the strength of diverging projections, distinct features may be independently adjusted while maintaining upstream coordination of the behavioural state. Further work will be needed to determine circuit mechanisms by which functional differentiation of these pathways originates, as well as how coordination ultimately occurs. Coordinated recruitment of the different populations of adBNST projection neurons could involve recurrent excitation (Supplementary Figs 16 and 17); indeed, *in vivo* multiunit recordings support the existence of recurrent excitation in the adBNST, as persistent activity was seen in 28% of recordings after

suppressed multiunit activity in adBNST. **e**, eNpHR3.0:BLA-adBNST mice were run in the EPM test for 20 min with alternating 1-min light-off and light-on epochs. **f**, Left, scatterplot of EPM scores in light-off and light-on conditions. Right, spatial firing maps illustrating change in EPM score of one single unit (red point in scatterplot) in response to yellow light, which decreased EPM score of most ($n = 28$ out of 38) units. **g**, Summary data across single units ($n = 38$): mean change in EPM score with inhibition of the adBNST projection. Notably, EPM scores even in light-on epochs were significantly higher than EPM scores generated from random simulated spikes ($P < 0.01$), indicating that even in light-on, BNST units could differentiate closed and open arms, although less robustly than in light-off. Data are mean \pm s.e.m. $*P < 0.05$; $***P < 0.001$. Statistics in Supplementary Information.

termination of BLA fibre stimulation (Supplementary Fig. 17a–c), and Ca^{2+} imaging in acute BNST slice revealed persistent activity in the adBNST after a single brief stimulus (Supplementary Fig. 17d–i).

Many complexities are involved in anxiety, including brain regions, hormonal changes, and physiological manifestations beyond those investigated here. For example, none of our manipulations altered heart rate (Supplementary Fig. 24), consistent with a previous report suggesting the BNST does not modulate this feature of anxiety¹⁷ and pointing to the need for further exploration of sympathetic pathways. Moreover, the anxious state may be parsed still further to delineate additional features, such as changes in exploratory drive or in novelty seeking, which could involve networks not explored here. It is likely that complex circuit structure and dynamics are required to assemble behavioural states in animals with highly diverse repertoires of internal states and adaptations to the environment.

METHODS SUMMARY

Virus-mediated gene expression. AAV5 viruses were packaged by the University of North Carolina Vector Core. Maps for the adeno-associated virus constructs are available at <http://www.optogenetics.org>. Viral stock (0.5 μl) was injected stereotactically into the BLA, BNST, lateral hypothalamus, parabrachial nucleus or VTA. **Anxiety assays and respiratory rate measurement.** Mice injected with viruses and implanted with guide cannulae or fibre-optics were subsequently tested in the EPM test, the OFT and the RTPP test. An EPM test session was 15-min long, consisting of 5-min light off-on-off-off epochs; the OFT was 20-min long, consisting of 5-min light off-on-off-on epochs. In the RTPP test, the subject could freely explore two chambers, and entry-to or exit-from one of the chambers turned optogenetic stimulation on or off, respectively. Behavioural data were automatically collected and analysed by BiObserve software. Respiratory rate was measured with a pulse oximeter from awake, behaving mice for 3 min. Yellow light was

delivered as constant illumination, whereas blue light was delivered as a train of 10-Hz, 5-ms pulses.

In vivo physiology. Custom-made microdrives containing eight stereotrodes surrounding a fibre-optic were implanted in the BNST, allowing for light delivery and recording of BNST neurons in awake, behaving animals. See Supplementary Methods for further details of analysis and computation of EPM scores.

Ex vivo electrophysiology. Acute slices were prepared for slice patch-clamp recordings. Whole-cell recordings were conducted from BNST neurons and blue light pulses at 10 Hz were delivered onto coronal sections via the microscope objective.

Statistics. All graphs and numerical values in the figures are presented as mean \pm s.e.m. See Supplementary Information.

Received 7 September 2012; accepted 18 February 2013.

Published online 20 March 2013.

- LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
- Davis, M. The role of the amygdala in fear and anxiety. *Annu. Rev. Neurosci.* **15**, 353–375 (1992).
- LeDoux, J. E., Iwata, J., Cicchetti, P. & Reis, D. J. Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *J. Neurosci.* **8**, 2517–2529 (1988).
- Viviani, D. *et al.* Oxytocin selectively gates fear responses through distinct outputs from the central amygdala. *Science* **333**, 104–107 (2011).
- Kessler, R. C. *et al.* Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **62**, 593–602 (2005).
- Singewald, N., Salchner, P. & Sharp, T. Induction of c-Fos expression in specific areas of the fear circuitry in rat forebrain by anxiogenic drugs. *Biol. Psychiatry* **53**, 275–283 (2003).
- Dong, H. W., Petrovich, G. D. & Swanson, L. W. Topography of projections from amygdala to bed nuclei of the stria terminalis. *Brain Res. Brain Res. Rev.* **38**, 192–246 (2001).
- Dabrowska, J. *et al.* Neuroanatomical evidence for reciprocal regulation of the corticotrophin-releasing factor and oxytocin systems in the hypothalamus and the bed nucleus of the stria terminalis of the rat: Implications for balancing stress and affect. *Psychoneuroendocrinology* **36**, 1312–1326 (2011).
- Walker, D. L., Miles, L. A. & Davis, M. Selective participation of the bed nucleus of the stria terminalis and CRF in sustained anxiety-like versus phasic fear-like responses. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **33**, 1291–1308 (2009).
- Duvarci, S., Bauer, E. P. & Paré, D. The bed nucleus of the stria terminalis mediates inter-individual variations in anxiety and fear. *J. Neurosci.* **29**, 10357–10361 (2009).
- Yassa, M. A., Hazlett, R. L., Stark, C. E. L. & Hoehn-Saric, R. Functional MRI of the amygdala and bed nucleus of the stria terminalis during conditions of uncertainty in generalized anxiety disorder. *J. Psychiatr. Res.* **46**, 1045–1052 (2012).
- Straube, T., Mentzel, H.-J. & Miltner, W. H. R. Waiting for spiders: brain activation during anticipatory anxiety in spider phobics. *Neuroimage* **37**, 1427–1436 (2007).
- Carola, V., D'Olimpio, F., Brunamonti, E., Mangia, F. & Renzi, P. Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. *Behav. Brain Res.* **134**, 49–57 (2002).
- Gradinaru, V. *et al.* Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154–165 (2010).
- Martin, B. The assessment of anxiety by physiological behavioral measures. *Psychol. Bull.* **58**, 234–255 (1961).
- Suess, W. M., Alexander, A. B., Smith, D. D., Sweeney, H. W. & Marion, R. J. The effects of psychological stress on respiration: a preliminary study of anxiety and hyperventilation. *Psychophysiology* **17**, 535–540 (1980).
- Terreberry, R. R., Oguri, M. & Harper, R. M. State-dependent respiratory and cardiac relationships with neuronal discharge in the bed nucleus of the stria terminalis. *Sleep* **18**, 139–144 (1995).
- Dong, H.-W. & Swanson, L. W. Organization of axonal projections from the anterolateral area of the bed nuclei of the stria terminalis. *J. Comp. Neurol.* **468**, 277–298 (2004).
- Dunn, J. D. & Williams, T. J. Cardiovascular responses to electrical stimulation of the bed nucleus of the stria terminalis. *J. Comp. Neurol.* **352**, 227–234 (1995).
- Dunn, J. D. Plasma corticosterone responses to electrical stimulation of the bed nucleus of the stria terminalis. *Brain Res.* **407**, 327–331 (1987).
- Woods, J. H., Katz, J. L. & Winger, G. Benzodiazepines: use, abuse, and consequences. *Pharmacol. Rev.* **44**, 151–347 (1992).
- Koob, G. F. A role for brain stress systems in addiction. *Neuron* **59**, 11–34 (2008).
- Dong, H.-W. & Swanson, L. W. Projections from bed nuclei of the stria terminalis, anteromedial area: cerebral hemisphere integration of neuroendocrine, autonomic, and behavioral aspects of energy balance. *J. Comp. Neurol.* **494**, 142–178 (2006).
- Dong, H. W., Petrovich, G. D., Watts, A. G. & Swanson, L. W. Basic organization of projections from the oval and fusiform nuclei of the bed nuclei of the stria terminalis in adult rat brain. *J. Comp. Neurol.* **436**, 430–455 (2001).
- Hakvoort-Schwerdtfeger, R. M. & Menard, J. L. The lateral hypothalamus and anterior hypothalamic nucleus differentially contribute to rats' defensive responses in the elevated plus-maze and shock-probe burying tests. *Physiol. Behav.* **93**, 697–705 (2008).
- Chamberlin, N. L. & Saper, C. B. Topographic organization of respiratory responses to glutamate microstimulation of the parabrachial nucleus in the rat. *J. Neurosci.* **14**, 6500–6510 (1994).
- Sartor, G. C. & Aston-Jones, G. Regulation of the ventral tegmental area by the bed nucleus of the stria terminalis is required for expression of cocaine preference. *Eur. J. Neurosci.* **36**, 3549–3558 (2012).
- Georges, F. & Aston-Jones, G. Activation of ventral tegmental area cells by the bed nucleus of the stria terminalis: a novel excitatory amino acid input to midbrain dopamine neurons. *J. Neurosci.* **22**, 5173–5187 (2002).
- Poulin, J.-F., Arbour, D., Laforest, S. & Drolet, G. Neuroanatomical characterization of endogenous opioids in the bed nucleus of the stria terminalis. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **33**, 1356–1365 (2009).
- Adhikari, A., Topiwala, M. A. & Gordon, J. A. Single units in the medial prefrontal cortex with anxiety-related firing patterns are preferentially influenced by ventral hippocampal activity. *Neuron* **71**, 898–910 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Davis, D. Walker, D. Paré, D. Rainnie, H. Shin, K. Thompson, P. Anikeeva, T. Davidson, I. Goshen, A. Andalman, L. Gunaydin, A. Bryant, C. Lee, J. Mirzabekov and the entire Deisseroth laboratory for discussions. Supported by a Samsung Scholarship (to S.-Y.K.), the US National Institute of Mental Health (NIMH; to R.C.M.), and a Berry Fellowship (to A.A.). K.D. and M.R.W. are NARSAD grant awardees, and K.D. was supported by the Wiegers Family Fund, the NIMH, the US National Institute on Drug Abuse (NIDA), the DARPA REPAIR Program, the Keck Foundation, the McKnight Foundation, the Gatsby Charitable Foundation, the Snyder Foundation, the Woo Foundation, the Tarlton Foundation, and the Albert Yu and Mary Bechman Foundation. All tools and methods are distributed and supported freely (<http://www.optogenetics.org>).

Author Contributions S.-Y.K., A.A. and K.D. designed the study, interpreted results and wrote the paper. S.-Y.K. coordinated the experiments. S.-Y.K., A.A., S.Y.L., C.S.M., M.R.W. and K.M.T. performed optogenetic behaviour and electrophysiology experiments. S.-Y.K., M.L., S.P. and J.M. performed immunohistochemistry. J.H.M. and C.K.K. performed calcium imaging. B.K.L., R.C.M. and R.N. provided viruses. All authors contributed to editing. K.D. supervised all aspects of the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.D. (deissero@stanford.edu).

Distinct extended amygdala circuits for divergent motivational states

Joshua H. Jennings^{1,2*}, Dennis R. Sparta^{1,3*}, Alice M. Stamatakis^{1,2}, Randall L. Ung¹, Kristen E. Pleil^{3,4}, Thomas L. Kash^{2,3,4,5} & Garret D. Stuber^{1,2,3,5,6}

The co-morbidity of anxiety and dysfunctional reward processing in illnesses such as addiction¹ and depression² suggests that common neural circuitry contributes to these disparate neuropsychiatric symptoms. The extended amygdala, including the bed nucleus of the stria terminalis (BNST), modulates fear and anxiety^{3,4}, but also projects to the ventral tegmental area (VTA)^{5,6}, a region implicated in reward and aversion^{7–13}, thus providing a candidate neural substrate for integrating diverse emotional states. However, the precise functional connectivity between distinct BNST projection neurons and their postsynaptic targets in the VTA, as well as the role of this circuit in controlling motivational states, have not been described. Here we record and manipulate the activity of genetically and neurochemically identified VTA-projecting BNST neurons in freely behaving mice. Collectively, aversive stimuli exposure produced heterogeneous firing patterns in VTA-projecting BNST neurons. By contrast, *in vivo* optically identified glutamatergic projection neurons displayed a net enhancement of activity to aversive stimuli, whereas the firing rate of identified GABAergic (γ -aminobutyric acid-containing) projection neurons was suppressed. Channelrhodopsin-2-assisted circuit mapping revealed that both BNST glutamatergic and GABAergic projections preferentially innervate postsynaptic non-dopaminergic VTA neurons, thus providing a mechanistic framework for *in vivo* circuit perturbations. *In vivo* photostimulation of BNST glutamatergic projections resulted in aversive and anxiogenic behavioural phenotypes. Conversely, activation of BNST GABAergic projections produced rewarding and anxiolytic phenotypes, which were also recapitulated by direct inhibition of VTA GABAergic neurons. These data demonstrate that functionally opposing BNST to VTA circuits regulate rewarding and aversive motivational states, and may serve as a crucial circuit node for bidirectionally normalizing maladaptive behaviours.

The ventral BNST (vBNST) is a heterogeneous structure¹⁴ that innervates the VTA^{5,15–18}, and aversive and rewarding stimuli activate a subset of these vBNST projection neurons^{19–21}. To identify and record the activity of vBNST–VTA neurons using antidromic photostimulation *in vivo*, we targeted channelrhodopsin-2 fused to enhanced yellow fluorescent protein (ChR2–eYFP)²² under the control of a CaMKII α promoter to the vBNST of adult mice. After 4–6 weeks, ChR2–eYFP was observed in vBNST cell bodies and projection fibres that innervate the VTA (Fig. 1a). Under anaesthesia, optical fibres for antidromic photostimulation were positioned above the VTA, whereas recording electrodes and optical fibres for orthodromic photostimulation were positioned in the vBNST (Fig. 1b). We recorded from vBNST units that showed reliable spiking to both orthodromic and antidromic photostimulation. By systematically decreasing the interval between orthodromic and antidromic photostimulation, the fidelity of antidromic spikes was significantly attenuated (Fig. 1c, d), demonstrating spike collision²³. In addition, antidromic spike latencies were significantly

greater and showed less variability than orthodromic spikes (Fig. 1e, f), and antidromic spike fidelity was significantly greater than orthodromic spike fidelity to 40-Hz photostimulation (Fig. 1g). Thus, photostimulation of vBNST–VTA projections results in antidromic spiking that is reliably distinguishable from putative trans-synaptic circuit activation.

To examine the neurophysiological dynamics of identified vBNST–VTA neurons in behaving mice, we implanted 16-channel multielectrode arrays in the vBNST, as well as optical fibres above the VTA for antidromic identification of neurons²⁴ (Fig. 1h and Supplementary Fig. 1). Delivery of single 5-ms, 473-nm light pulses to the VTA resulted in time-locked firing in many vBNST neurons. Photostimulation of vBNST–VTA fibres resulted in a bimodal firing pattern in vBNST neurons due to distinguishable antidromic and polysynaptic activity (Supplementary Fig. 2 and Methods). Principle component and correlation analysis comparing waveform shapes demonstrated that spontaneous waveforms were highly correlated with light-evoked waveforms⁷ (average $r = 0.950 \pm 0.008$ (mean \pm s.e.m.); Supplementary Fig. 2 and Supplementary Table 1). Light-evoked spike latencies showed that a subset of recorded units consistently displayed time-locked spiking on 11.21 ± 0.68 out of 20 trials (56%), with a mean latency of 7.31 ± 0.32 ms (Fig. 1i, j), comparable with our anaesthetized recording data (Fig. 1e) and a previous study using electrical antidromic stimulation of BNST projections in rodents²⁵. Accordingly, neurons that were identified as antidromic-responsive displayed a spike fidelity of $81 \pm 15\%$ in response to 20-Hz photostimulation (Supplementary Fig. 2). Using these criteria (Methods), we identified 53 out of 137 units as vBNST–VTA projection neurons.

vBNST neurons display heterogeneous responses to aversive stimuli exposure^{19,20}. Thus, we classified the firing patterns of identified vBNST–VTA neurons in response to unpredictable foot shocks and associated contextual cues (Methods). Identified vBNST–VTA neurons segregated into three functionally distinct classes based on changes in their normalized firing rates throughout the foot-shock session (Supplementary Fig. 3), demonstrating that vBNST–VTA neurons differentially encode information related to aversive stimuli and their associated contextual cues.

Electrical stimulation of the BNST produces both excitatory and inhibitory responses in VTA neurons *in vivo*⁶, suggesting that distinct subcircuits may exist. Mice were injected with adeno-associated viral vector (AAV5) expressing ChR2–eYFP under the control of the CaMKII α promoter (AAV5 CaMKII α ::ChR2–eYFP) to nonspecifically target vBNST–VTA projection neurons (CaMKII α ^{vBNST–VTA}::ChR2). Whole-cell recordings in brain slices revealed that photostimulation of the CaMKII α ^{vBNST–VTA} pathway produced both glutamatergic and GABAergic currents in VTA neurons (Supplementary Fig. 4), demonstrating that neurochemically distinct vBNST neurons project to the VTA.

¹Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ²Neurobiology Curriculum, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ³Bowles Center for Alcohol Studies, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ⁴Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ⁵Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ⁶Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

*These authors contributed equally to this work.

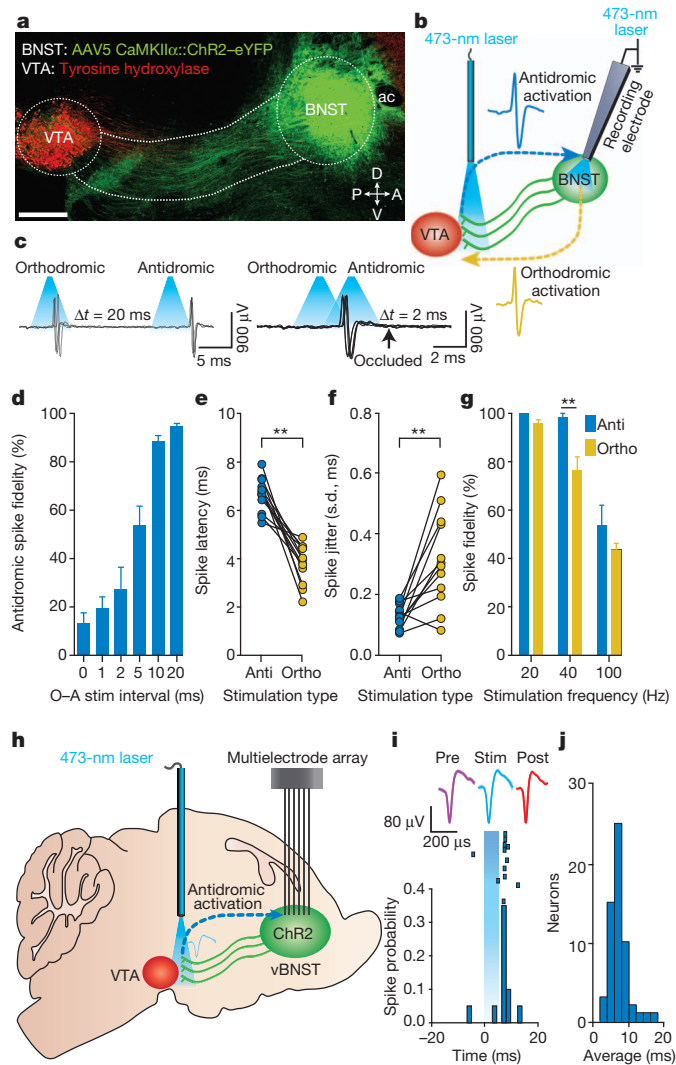


Figure 1 | Optogenetic identification of vBNST-VTA projection neurons. **a**, Sagittal image showing the vBNST-VTA projection. A, anterior; ac, anterior commissure; D, dorsal; P, posterior; V, ventral. Scale bar, 500 μ m. **b**, Optogenetic collision test. **c**, Example traces from a single CaMKII α ^{vBNST-VTA} unit demonstrating antidromic-orthodromic spike collision. **d**, Significant reduction in the percentage of antidromic spike fidelity at short antidromic-orthodromic (A-O) photostimulation intervals ($F_{5,65} = 48.63$, $P < 0.0001$; $n = 12$ units). **e**, Antidromic (anti) spike latencies were significantly greater than orthodromic (ortho) latencies ($P < 0.0001$; $n = 12$ units). **f**, Antidromic-initiated spikes displayed significantly greater latency stability than orthodromic-activated spikes ($P < 0.001$; $n = 12$ units). **g**, Antidromic spikes responded more reliably to 40-Hz photostimulation than orthodromic spikes ($F_{2,18} = 11.2$, $P = 0.003$, $n = 4$ units). **h**, Optogenetic identification of vBNST-VTA projection neurons in behaving mice. **i**, Representative peristimulus time histogram and raster of a single unit time-locked to 5-ms antidromic photostimulation. **j**, Mean first-spike latencies after antidromic photostimulation for all identified CaMKII α ^{vBNST-VTA} projection neurons ($n = 53$ units, $n = 7$ mice). Data are mean \pm s.e.m. $^{**}P < 0.01$ (Student's *t*-tests and Bonferroni post-hoc comparisons, where applicable).

We next dissected the functional connectivity between distinct glutamatergic and GABAergic vBNST-VTA neurons and their genetically defined postsynaptic targets within the VTA. Injection of a Cre-inducible viral construct coding for ChR2-eYFP into the vBNST in *Vglut2*-ires-cre or *Vgat*-ires-cre mouse lines (which express Cre recombinase from the endogenous *Vglut2* (also known as *Slc17a6*) or *Vgat* (*Slc32a1*) promoters, respectively)²⁶ resulted in robust expression in the vBNST as well as in fibres originating from these neurons that innervated the VTA (Fig. 2a, b). Whole-cell recordings from VTA neurons revealed that photostimulation of ChR2-containing fibres originating from vBNST

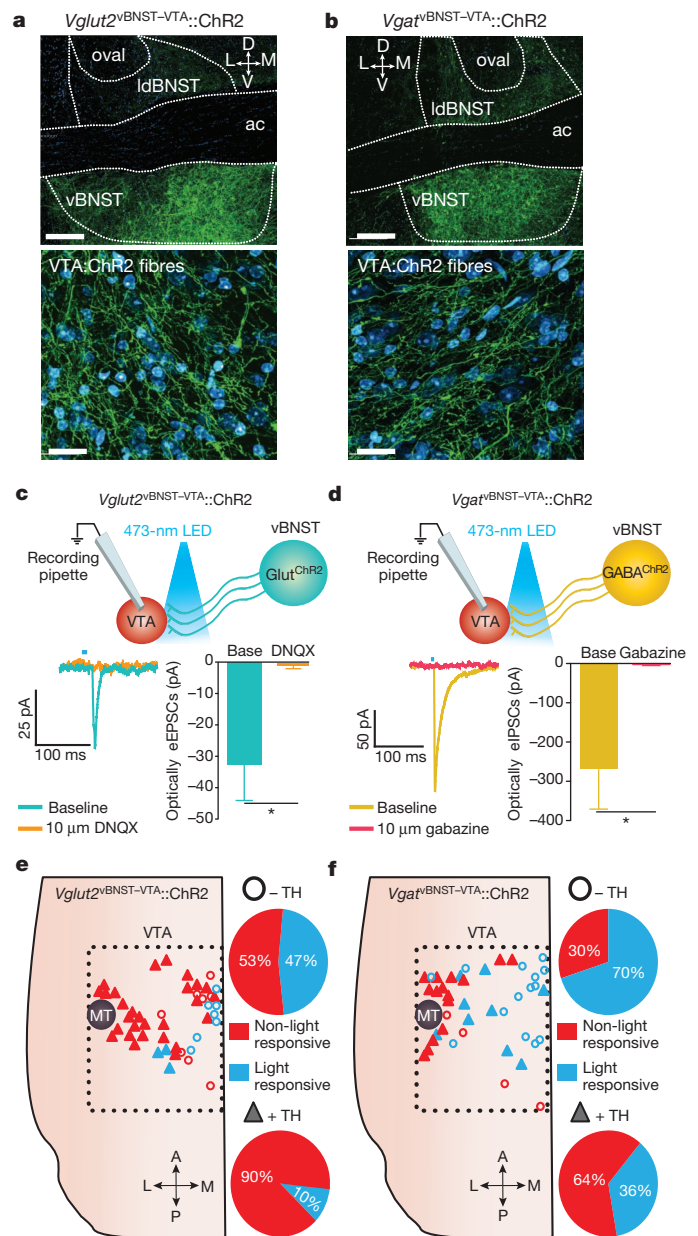


Figure 2 | Excitatory and inhibitory synapses onto non-dopaminergic VTA neurons from neurochemically distinct vBNST neurons. **a**, **b**, ChR2-eYFP (green) in the vBNST (top) and fibres in the VTA (bottom) in *Vglut2*-ires-cre (**a**) and *Vgat*-ires-cre (**b**) mice. Cyan denotes fluorescent Nissl stain. Scale bars, 200 μ m (top) and 20 μ m (bottom). L, lateral; ldBNST, lateral-dorsal BNST; M, medial; oval, oval nucleus BNST. **c**, Optically evoked excitatory postsynaptic currents (eEPSCs) recorded in VTA neurons after *Vglut2*^{vBNST-VTA::ChR2} stimulation before and after application of the glutamate receptor antagonist 6,7-dinitroquinoxaline-2,3-dione (DNQX) (bottom) ($n = 4$ cells, $P = 0.0307$). LED, light-emitting diode. **d**, Optically evoked inhibitory postsynaptic currents (eIPSCs) recorded in VTA neurons after *Vgat*^{vBNST-VTA::ChR2} stimulation before and after application of the GABA_A receptor antagonist gabazine (bottom) ($n = 4$ cells, $P = 0.0378$). **e**, **f**, Location of light-responsive and non-light-responsive dopaminergic and non-dopaminergic neurons in horizontal VTA slices after photostimulation of *Vglut2*^{vBNST-VTA::ChR2} (**e**) and *Vgat*^{vBNST-VTA::ChR2} (**f**) projections. TH, tyrosine hydroxylase. Data are mean \pm s.e.m. $^{*}P < 0.05$ (Student's *t*-test for paired samples).

neurons expressing *Vglut2* (*Vglut2*^{vBNST-VTA::ChR2}) or *Vgat* (*Vgat*^{vBNST-VTA::ChR2}) produced excitatory or inhibitory postsynaptic currents, respectively (Fig. 2c, d and Supplementary Fig. 5). *Vglut2*^{vBNST-VTA} and *Vgat*^{vBNST-VTA} terminals formed functional

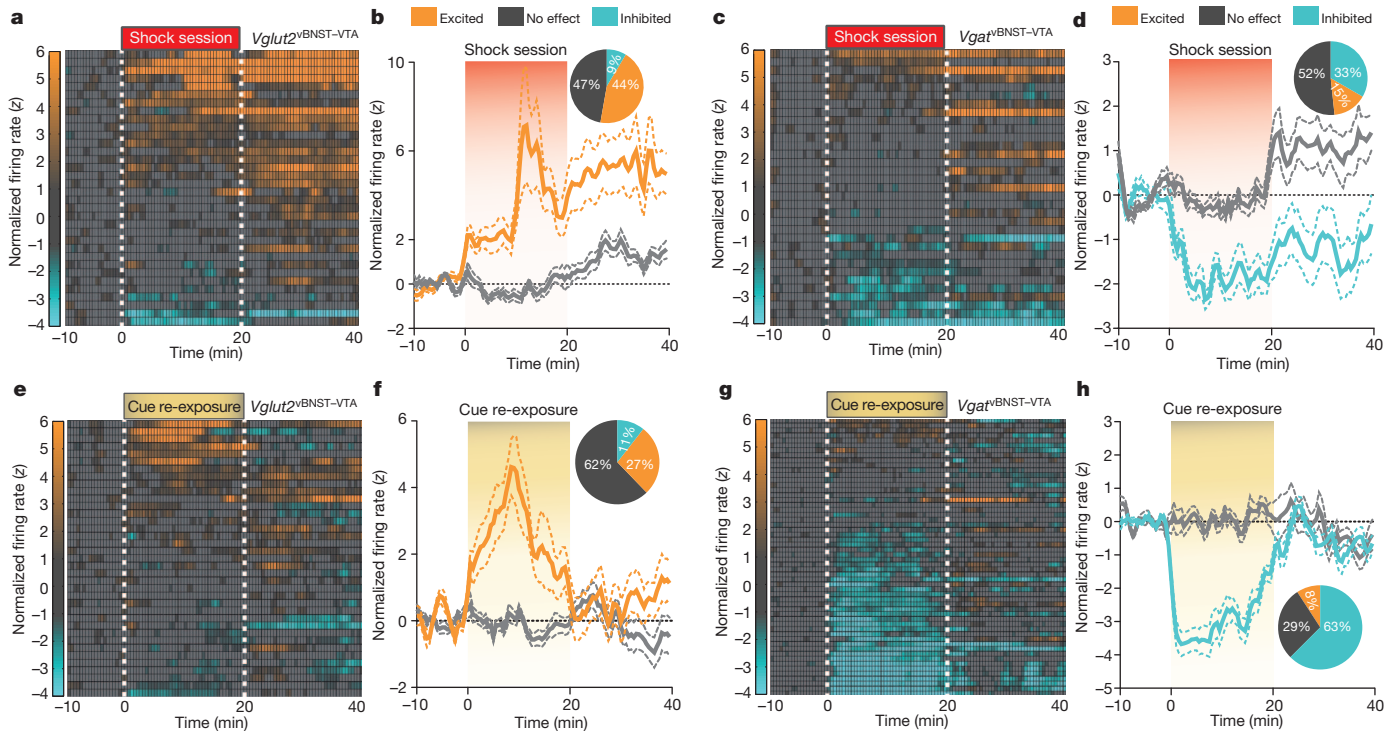


Figure 3 | *Vglut2*^{vBNST-VTA} and *Vgat*^{vBNST-VTA} projection neurons display distinct firing patterns in response to foot-shock and shock-associated contextual cues. **a**, Colour-coded normalized firing rates for all identified *Vglut2*^{vBNST-VTA} neurons in response to the first foot-shock session. **b**, Average normalized firing rate of classified shock-excited *Vglut2*^{vBNST-VTA} neurons is significantly altered compared to classified no-effect neurons during and after the foot-shock session ($F_{99,2900} = 3.13$, $P < 0.0001$, $n = 34$ units, $n = 7$ mice). Inset, percentages of classified neurons. **c**, Colour-coded normalized firing rates for all identified *Vgat*^{vBNST-VTA} neurons in response to the first foot-shock session. **d**, Average normalized firing rate of classified shock-inhibited *Vgat*^{vBNST-VTA} neurons is significantly altered compared to classified no-effect

neurons during and after the foot-shock session ($F_{99,2600} = 2.66$, $P < 0.0001$, $n = 33$ units, $n = 5$ mice). **e**, Colour-coded normalized firing rates of identified *Vglut2*^{vBNST-VTA} neurons in response to cue re-exposure. **f**, Average normalized firing rate of classified cue-excited *Vglut2*^{vBNST-VTA} neurons is significantly altered compared to classified no-effect neurons during and after cue re-exposure ($F_{99,3100} = 5.135$, $P < 0.0001$, $n = 37$ units, $n = 4$ mice). **g**, Colour-coded normalized firing rates of *Vgat*^{vBNST-VTA} neurons in response to cue re-exposure. **h**, Average normalized firing rate of classified cue-inhibited *Vgat*^{vBNST-VTA} neurons is significantly altered compared to classified no-effect neurons during and after cue re-exposure ($F_{99,4900} = 8.285$, $P < 0.0001$, $n = 56$ units, $n = 4$ mice).

synapses primarily onto non-dopaminergic and medially located dopaminergic neurons, which have been implicated in responding to aversive stimuli^{7,9,11,13} (Fig. 2e, f and Supplementary Figs 6 and 7 and Methods). These data provide a circuit blueprint by which vBNST subcircuits interact with VTA-reward circuitry.

We next explored whether glutamatergic or GABAergic subpopulations of vBNST-VTA neurons differentially respond to foot-shock sessions and associated contextual cues. Using optical antidromic activation *in vivo*, we identified 34 *Vglut2*^{vBNST-VTA::ChR2}-expressing neurons out of 145 recorded neurons (Supplementary Figs 1 and 8 and Supplementary Table 1). Although all projection neurons displayed heterogeneous firing patterns (Supplementary Fig. 3), identified *Vglut2*^{vBNST-VTA} projection neurons exhibited a net enhancement of firing during the aversive event (Fig. 3a, b). By contrast, 33 identified *Vgat*^{vBNST-VTA::ChR2}-expressing neurons out of 77 total neurons principally exhibited reduced firing during the aversive event (Fig. 3c, d, Supplementary Figs 1 and 8 and Supplementary Table 1). In addition, 1 week after five consecutive daily foot-shock sessions, re-exposure to shock-associated contextual cues alone resulted in a net enhancement of *Vglut2*^{vBNST-VTA::ChR2} neuronal activity (Fig. 3e, f and Supplementary Fig. 9), whereas the activity of *Vgat*^{vBNST-VTA::ChR2} neurons was largely suppressed (Fig. 3g, h and Supplementary Fig. 9). Collectively, exposure to the aversive event or associated cues alone enhanced the firing of *Vglut2*^{vBNST-VTA} neurons, while simultaneously suppressing the activity of *Vgat*^{vBNST-VTA} neurons.

Because aversive stimuli enhanced the activity of *Vglut2*^{vBNST-VTA} neurons (Fig. 3a, b, e, f), which can excite non-dopaminergic VTA neurons (Fig. 2e), we next explored the behavioural consequences of

selectively activating this projection in behaving mice. We tested mice in a real-time place preference (RTPP) model to assay the effects of photostimulation of the *Vglut2*^{vBNST-VTA} pathway on motivational valence. Photostimulation of *Vglut2*^{vBNST-VTA::ChR2} mice resulted in a significant avoidance of a stimulation-paired chamber (Fig. 4a, b and Supplementary Figs 10 and 11). Activation of this pathway also reduced active reward seeking (Supplementary Fig. 11). The aversive effects of this stimulation were dependent on local VTA glutamatergic signalling as infusions of an ionotropic glutamate receptor antagonist cocktail abolished the aversive phenotype induced by *Vglut2*^{vBNST-VTA} activation (Fig. 4b and Supplementary Figs 12 and 13). In addition, inescapable activation of this pathway for 20 min in an open field resulted in significantly less centre and more corner time in *Vglut2*^{vBNST-VTA::ChR2} mice in the 10 min after stimulation offset compared to controls, suggesting that enhanced activity in the *Vglut2*^{vBNST-VTA} pathway contributes to anxiety-like behaviour (Fig. 4c and Supplementary Fig. 11).

In contrast to the aversive consequences of stimulating the *Vglut2*^{vBNST-VTA} pathway, 20-Hz photostimulation in *Vgat*^{vBNST-VTA::ChR2} mice resulted in a significant place preference (Fig. 5a, b and Supplementary Figs 10 and 14). VTA infusions of a GABA_A receptor antagonist prevented the *Vgat*^{vBNST-VTA}-mediated place preference compared to saline injections (Fig. 5b and Supplementary Figs 12 and 13). To determine whether *in vivo* optogenetic activation of the *Vgat*^{vBNST-VTA} pathway produces active reward seeking, we tested whether these mice would nose poke to receive photostimulation²⁷. *Vgat*^{vBNST-VTA::ChR2} mice readily nose poked to receive photostimulation (Fig. 5c and Supplementary Fig. 14). Together, these data suggest

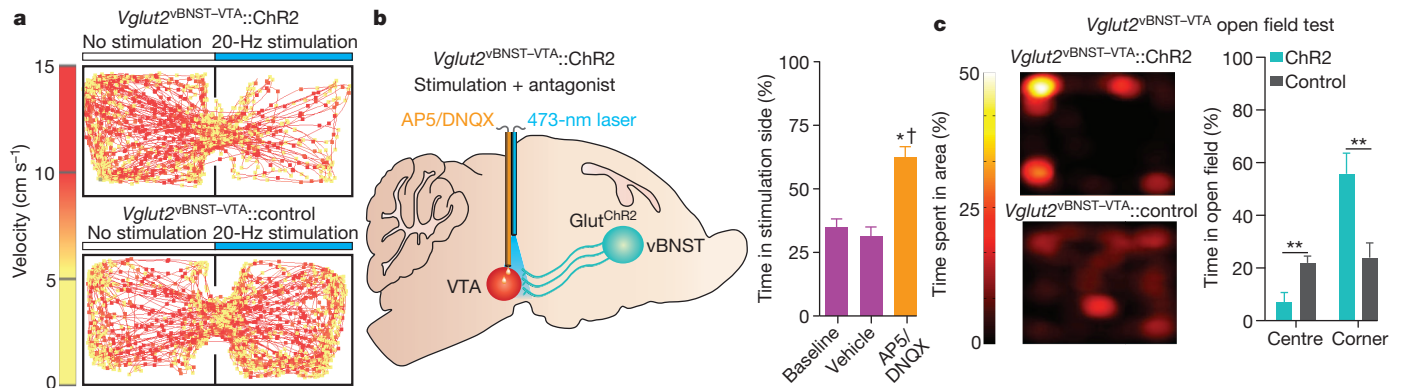


Figure 4 | Photostimulation of the *Vglut2*^{vBNST-VTA} pathway promotes aversion and anxiety. **a**, Representative RTTP tracks from *Vglut2*^{vBNST-VTA}::ChR2 (top) and control (bottom) mice. **b**, Intra-VTA infusions of a glutamate antagonist cocktail (D(-)-2-amino-5-phosphonovaleric acid (AP5) plus DNQX), followed by *Vglut2*^{vBNST-VTA}::ChR2 stimulation during blocked aversion in the RTTP test ($F_{3,15} = 12.811$, $P < 0.001$, $n = 6$ mice). **c**, Representative heat maps displaying average time

spent in an open field for 10 min after stimulation from *Vglut2*^{vBNST-VTA}::ChR2 (top) and *Vglut2*^{vBNST-VTA}::control (bottom) mice. *Vglut2*^{vBNST-VTA}::ChR2 mice spent significantly more time in the corners ($P = 0.008$) and less time in the centre ($P = 0.007$) of an open field immediately after constant 20-Hz stimulation than *Vglut2*^{vBNST-VTA}::control mice ($n = 6$ mice per group). Data are mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$. Dagger symbol denotes significance compared to all manipulations.

that photostimulation of the *Vgat*^{vBNST-VTA} pathway promotes reward-related behaviours.

Because the *Vgat*^{vBNST-VTA} projection preferentially innervates non-dopaminergic VTA neurons (Fig. 2f), we considered VTA GABAergic

neurons as the likely postsynaptic target. VTA GABAergic neuronal inhibition via halorhodopsin activation with enhanced *Natromonas pharaonis* halorhodopsin (eNpHR3.0) (Supplementary Figs 15 and 16) also produced reward-related phenotypes (*Vgat*^{VTA}::eNpHR3.0;

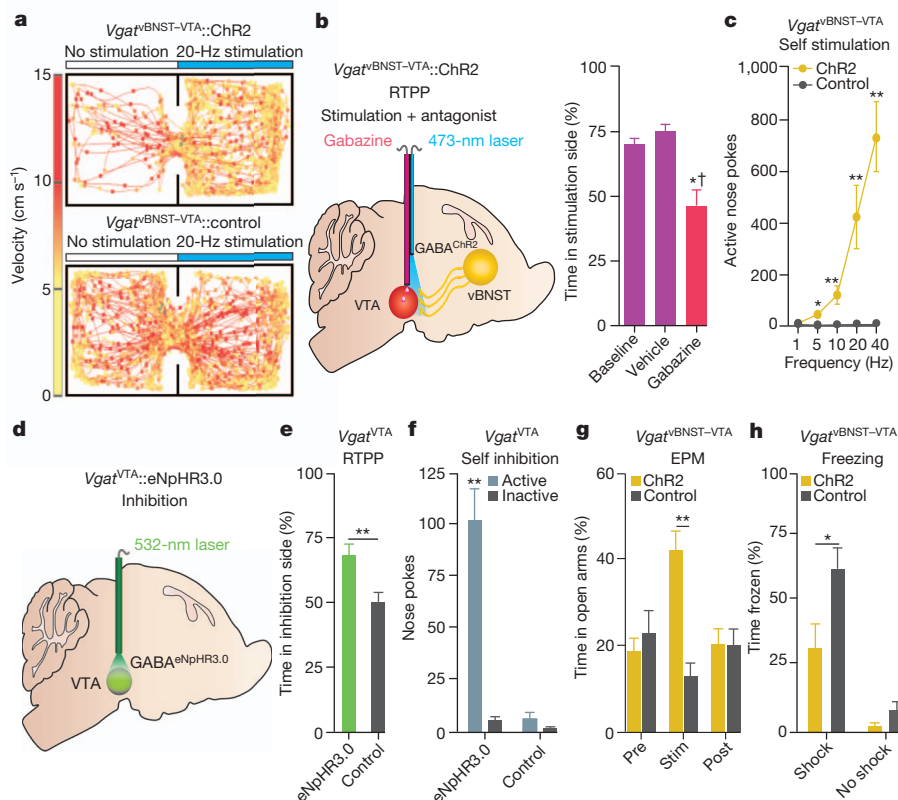


Figure 5 | Photostimulation of the *Vgat*^{vBNST-VTA} pathway and inhibition of *Vgat*^{VTA} neurons produces reward-related behaviours and attenuates anxiety. **a**, Representative RTTP tracks from *Vgat*^{vBNST-VTA}::ChR2 (top) and control (bottom) mice. **b**, Intra-VTA infusions of the GABA_A antagonist gabazine followed by *Vgat*^{vBNST-VTA}::ChR2 stimulation abolished place preference ($F_{3,15} = 13.718$, $P < 0.001$, $n = 6$ mice). **c**, *Vgat*^{vBNST-VTA}::ChR2 mice made significantly more nose pokes to obtain photostimulation than controls ($F_{4,36} = 12.42$, $P < 0.001$, $n = 5-7$ mice per group). **d**, Schematic detailing *Vgat*^{VTA}::eNpHR3.0 inhibition during behavioural experiments. **e**, *Vgat*^{VTA}::eNpHR3.0 mice spent significantly more time in the inhibition-paired

side than controls ($P = 0.01$, $n = 6$ mice per group). **f**, *Vgat*^{VTA}::eNpHR3.0 mice made significantly more nose pokes to obtain photoinhibition than controls ($P < 0.001$, $n = 5$ mice per group). **g**, *Vgat*^{vBNST-VTA}::ChR2 mice spent significantly more time in the EPM open arms than controls during the 5-min photostimulation epoch ($F_{2,24} = 14.648$, $P < 0.001$, $n = 7$ mice per group). **h**, After concurrent photostimulation during the foot-shock session, *Vgat*^{vBNST-VTA}::ChR2 mice ($n = 6-7$) spent significantly less time frozen than controls ($F_{1,22} = 37.992$, $P < 0.001$). Data are mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$. Dagger symbol denotes significance compared to all manipulations.

Fig. 5d–f). Together, these results show that reward-related responses to $Vgat^{vBNST-VTA}$ activation are recapitulated by directly inhibiting $Vgat^{VTA}$ neurons, thus providing a circuit mechanism for the $Vgat^{vBNST-VTA}$ pathway to regulate motivated behaviour.

As the BNST regulates the expression of fear and anxiety-related behavioural phenotypes^{3,28,29}, we also sought to establish a role for the $Vgat^{vBNST-VTA}$ pathway in these negative motivational states. Photostimulation of the $Vgat^{vBNST-VTA}$ pathway and direct inhibition of $Vgat^{VTA}$ neurons significantly increased the time spent in the open arms of an elevated-plus maze (EPM) test, indicative of anxiolysis (Fig. 5g and Supplementary Fig. 17). These coinciding observations suggest that $Vgat^{vBNST-VTA}$ and $Vgat^{VTA}$ neurons act as crucial circuit nodes for moderating the expression of anxiety.

Given that $Vgat^{vBNST-VTA}$ neurons are largely inhibited by aversive stimuli (Fig. 3c, d, g, h), we examined whether concurrent activation of the $Vgat^{vBNST-VTA}$ projection during an unpredictable foot-shock session could alleviate the subsequent development of anxiety-like behaviour. Immediately after termination of the foot-shock session and cessation of $Vgat^{vBNST-VTA}::ChR2$ stimulation, we measured the acute freezing response while mice were still in the shock-associated context, as well as their behaviour in the EPM test 3 h later (Supplementary Fig. 18 and Methods). $Vgat^{vBNST-VTA}::ChR2$ mice spent significantly less time frozen (Fig. 5h), as well as significantly more open-arm time and entries in the EPM test relative to controls (Supplementary Fig. 18). Taken together, these data suggest that enhancing activity of the $Vgat^{vBNST-VTA}$ pathway during aversive stimuli exposure has anxiety-buffering properties. Although the canonical view of BNST function proposes a dominant role of this structure in promoting anxiety states^{3,4,30}, the cellular and functional complexity described here (Supplementary Fig. 19) illustrates that particular BNST circuit elements orchestrate divergent aspects of emotional and motivational processing.

METHODS SUMMARY

All procedures were conducted in accordance with the Guide for the Care and Use of Laboratory Animals, as adopted by the National Institutes of Health, and with approval of the Institutional Animal Care and Use Committee at the University of North Carolina and described in detail in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 7 September 2012; accepted 25 February 2013.

Published online 20 March 2013.

- Koob, G. F. & Le Moal, M. Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology* **24**, 97–129 (2001).
- Nestler, E. J. & Carlezon, W. A. Jr. The mesolimbic dopamine reward circuit in depression. *Biol. Psychiatry* **59**, 1151–1159 (2006).
- Davis, M., Walker, D. L., Miles, L. & Grillon, C. Phasic vs sustained fear in rats and humans: role of the extended amygdala in fear vs anxiety. *Neuropsychopharmacology* **35**, 105–135 (2010).
- Walker, D. L. & Davis, M. Role of the extended amygdala in short-duration versus sustained fear: a tribute to Dr. Lennart Heimer. *Brain Struct. Funct.* **213**, 29–42 (2008).
- Geisler, S. & Zahm, D. S. Afferents of the ventral tegmental area in the rat-anatomical substratum for integrative functions. *J. Comp. Neurol.* **490**, 270–294 (2005).
- Georges, F. & Aston-Jones, G. Potent regulation of midbrain dopamine neurons by the bed nucleus of the stria terminalis. *J. Neurosci.* **21**, RC160 (2001).
- Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
- Fields, H. L., Hjelmstad, G. O., Margolis, E. B. & Nicola, S. M. Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. *Annu. Rev. Neurosci.* **30**, 289–316 (2007).

- Lammel, S. *et al.* Input-specific control of reward and aversion in the ventral tegmental area. *Nature* **491**, 212–217 (2012).
- Stamatakis, A. M. & Stuber, G. D. Activation of lateral habenula inputs to the ventral midbrain promotes behavioral avoidance. *Nature Neurosci.* **15**, 1105–1107 (2012).
- Tan, K. R. *et al.* GABA neurons of the VTA drive conditioned place aversion. *Neuron* **73**, 1173–1183 (2012).
- Tye, K. M. *et al.* Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature* **493**, 537–541 (2012).
- van Zessen, R., Phillips, J. L., Budygin, E. A. & Stuber, G. D. Activation of VTA GABA neurons disrupts reward consumption. *Neuron* **73**, 1184–1194 (2012).
- Hammack, S. E., Mania, I. & Rainnie, D. G. Differential expression of intrinsic membrane currents in defined cell types of the anterolateral bed nucleus of the stria terminalis. *J. Neurophysiol.* **98**, 638–656 (2007).
- Dong, H. W. & Swanson, L. W. Organization of axonal projections from the anterolateral area of the bed nuclei of the stria terminalis. *J. Comp. Neurol.* **468**, 277–298 (2004).
- Dumont, E. C. & Williams, J. T. Noradrenergic triggers GABA inhibition of bed nucleus of the stria terminalis neurons projecting to the ventral tegmental area. *J. Neurosci.* **24**, 8198–8204 (2004).
- Jalabert, M., Aston-Jones, G., Herzog, E., Manzoni, O. & Georges, F. Role of the bed nucleus of the stria terminalis in the control of ventral tegmental area dopamine neurons. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **33**, 1336–1346 (2009).
- Kudo, T. *et al.* Three types of neurochemical projection from the bed nucleus of the stria terminalis to the ventral tegmental area in adult mice. *J. Neurosci.* **32**, 18035–18046 (2012).
- Briand, L. A., Vassoler, F. M., Pierce, R. C., Valentino, R. J. & Blendy, J. A. Ventral tegmental afferents in stress-induced reinstatement: the role of cAMP response element-binding protein. *J. Neurosci.* **30**, 16149–16159 (2010).
- Christianson, J. P. *et al.* Safety signals mitigate the consequences of uncontrollable stress via a circuit involving the sensory insular cortex and bed nucleus of the stria terminalis. *Biol. Psychiatry* **70**, 458–464 (2011).
- Mahler, S. V. & Aston-Jones, G. S. Fos activation of selective afferents to ventral tegmental area during cue-induced reinstatement of cocaine seeking in rats. *J. Neurosci.* **32**, 13309–13325 (2012).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
- Fuller, J. H. & Schlag, J. D. Determination of antidromic excitation by the collision test: problems of interpretation. *Brain Res.* **112**, 283–298 (1976).
- Sparta, D. R. *et al.* Construction of implantable optical fibers for long-term optogenetic manipulation of neural circuits. *Nature Protocols* **7**, 12–23 (2012).
- Nagy, F. Z. & Pare, D. Timing of impulses from the central amygdala and bed nucleus of the stria terminalis to the brain stem. *J. Neurophysiol.* **100**, 3429–3436 (2008).
- Vong, L. *et al.* Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* **71**, 142–154 (2011).
- Stuber, G. D. *et al.* Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature* **475**, 377–380 (2011).
- Erb, S., Shaham, Y. & Stewart, J. Stress-induced relapse to drug seeking in the rat: role of the bed nucleus of the stria terminalis and amygdala. *Stress* **4**, 289–303 (2001).
- Poulos, A. M., Ponnusamy, R., Dong, H. W. & Fanselow, M. S. Compensation in the neural circuitry of fear conditioning awakens learning circuits in the bed nuclei of the stria terminalis. *Proc. Natl Acad. Sci. USA* **107**, 14881–14886 (2010).
- Phelps, E. A. & LeDoux, J. E. Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* **48**, 175–187 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Patel, J. Phillips and S. Maciver for assistance; V. Gukasyan and the UNC Neuroscience Center Microscopy Core (P30 NS045892), and members of the Stuber laboratory for discussion. We thank K. Deisseroth for viral constructs and the UNC Vector Core Facility for viral packaging. We thank B. Lowell and L. Vong for providing the $Vgat$ -ires-cre and $Vglut2$ -ires-cre mice. This study was supported by The Whitehall Foundation, The Foundation of Hope, National Institutes of Health grants DA029325 and DA032750 (to G.D.S.), AA018610 and AA007573 (to D.R.S.), NS007431 and DA034472 (to A.M.S.) and AA021043 (to K.P.), and the UNC NIAAA alcohol research center (AA011605).

Author Contributions D.R.S., J.H.J. and G.D.S. designed all experiments and wrote the manuscript. All authors collected, analysed and discussed the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.D.S. (gstuber@med.unc.edu).

METHODS

Experimental subjects and stereotactic surgery. All procedures were conducted in accordance with the Guide for the Care and Use of Laboratory Animals, as adopted by the National Institutes of Health, and with approval of the Institutional Animal Care and Use Committee at the University of North Carolina (UNC). Adult (25–30 g) male C57BL/6J mice (Jackson Laboratory), adult male *Vgat-ires-cre* mice and adult male *Vglut2-ires-cre* mice (see ref. 26 for further details on the *Vglut2-ires-cre* and *Vgat-ires-cre* mouse lines) were group-housed before surgery. All mice were maintained on a reverse 12-h light cycle (lights off at 7:00) with *ad libitum* access to food and water, unless described otherwise. Mice were anaesthetized with a ketamine (150 mg per kg of body weight) and xylazine (50 mg per kg) solution and placed into a stereotactic frame (Kopf Instruments). For all *in vivo* electrophysiology experiments, male mice were unilaterally injected with 0.5 μ l of purified and concentrated adeno-associated virus (AAV) ($\sim 10^{12}$ infectious units per ml, packaged by the UNC Vector Core Facility) into the vBNST using the following stereotactic coordinates: +0.14 mm to bregma, ± 0.9 mm lateral to midline, and -4.8 mm ventral to the skull surface. All viral constructs were packaged by the UNC Vector Core Facility at a final working concentration of $1 \times 10^{12} - 5 \times 10^{12}$ genome copies per millilitre.

For all *in vivo* electrophysiology experiments, mice were implanted with a 16-wire (4×4 configuration, wire diameter $\sim 30 \mu$ m) tungsten multielectrode array (Innovative Neurophysiology) aimed at the vBNST using the stereotactic coordinates stated above. For all *in vivo* electrophysiological and behavioural experiments, except for the VTA microinjection experiments, all mice were implanted with an optical fibre aimed at the VTA (see ref. 24 for further details), using the following stereotactic coordinates: -3.2 mm to bregma, ± 0.5 mm lateral to midline, and -4.69 mm ventral to the skull surface. For the VTA microinjection experiments, a 26-gauge steel tube cannula (McMasters-Carr) that terminated 0.5 mm above the tip of the optical fibre was epoxied to an optical fibre and unilaterally aimed at the VTA using the following stereotactic coordinates: -3.2 mm to bregma, ± 0.5 mm lateral to midline, and -4.69 mm ventral to skull surface. For photoinhibition of VTA-GABAergic neurons using eNpHR3.0, all mice were bilaterally implanted with an optical fibre at a 10° angle in the VTA using the following stereotactic coordinates: -3.2 mm to bregma, ± 1.1 mm lateral to midline, and -4.75 mm ventral to the skull surface. The time from virus injection to the start of the experiments was 4–6 weeks for all Chr2 terminal stimulation manipulations, and 3–4 weeks for cell body manipulations.

Histology, immunohistochemistry and microscopy. Mice were anaesthetized with pentobarbital, and transcardially perfused with PBS followed by 4% (w/v) paraformaldehyde in PBS. Forty-micrometre brain sections were subjected to immunohistochemical staining for neuronal cell bodies (NeuroTrace Invitrogen; 640 nm excitation/660 nm emission or 435 nm excitation/455 nm emission and/or tyrosine hydroxylase) (Pel Freeze; made in sheep, 1:500) (see refs 10, 13 for additional information). Brain sections were mounted, coverslips added, and z-stack and tiled images were captured on a Zeiss LSM 710 confocal microscope using a $\times 20$ or $\times 63$ objective. To determine optical fibre placement, tissue was imaged at $\times 10$ and $\times 20$ on an upright epi-fluorescent microscope.

***In vivo* anaesthetized electrophysiology.** C57BL/6J mice were bilaterally injected with 0.3 μ l of AAV5 CaMKII α ::Chr2-eYFP into the vBNST. Six weeks after virus injection, mice were anaesthetized with 0.5–1.0% isoflurane (Butler Schein) and were placed into a stereotaxic frame (Kopf Instruments). Body temperature was maintained at $\sim 37^\circ\text{C}$ with a homeothermic heating blanket (Harvard Apparatus). Tail pinches were administered frequently to monitor responses under anaesthesia. A reference electrode was fixed inside brain tissue, approximately 2 mm from both the vBNST and VTA. Extracellular neural activity was recorded using a glass recording electrode (5–10 M Ω ; and filled with 0.5 M NaCl). The recording electrode was lowered into the vBNST (+0.14 mm to bregma, ± 0.9 mm lateral to midline, and -4.8 mm ventral to the skull surface) by a motorized micromanipulator (Scientific). Recordings were amplified (Multiclamp 700B, Molecular Devices), high-pass filtered at 6 kHz and sampled at 10 kHz. Here, orthodromic photostimulation refers to action potentials initiated at the cell body, whereas antidromic photostimulation refers to backward propagating action potentials initiated at distal axonal fibres; both are independent of synaptic transmission.

For orthodromic activation, an optical fibre coupled to a solid state laser (473 nm) was fed through the side port of the electrode holder to terminate near the tip of the glass recording electrode, which allowed for delivery of ~ 5 mW light pulses into the vBNST. For antidromic activation, an optical fibre housed in a steel cannula and coupled to a separate solid-state laser (473 nm) was inserted into the VTA at a 16° angle (-3.2 mm to bregma, $+1.4$ mm lateral to midline, and -4.9 mm ventral to the skull surface), which delivered ~ 10 mW of light to the VTA. vBNST neurons were classified as antidromic-responsive if the following three criteria were met: (1) stable antidromic spike latency (< 0.2 ms); (2) ability to respond reliably to high-frequency photostimulation; and (3) collision between

orthodromic- and antidromic-evoked spikes. Each photostimulation parameter delivered a 5-ms light pulse to either vBNST cell bodies (orthodromic) or vBNST axons within the VTA (antidromic). To determine stable antidromic latencies, 5-ms light pulses were delivered to the VTA every 5 s for 20 trials. To confirm reliable antidromic spike fidelity, 20-, 40- and 100-Hz train pulses of light were delivered to the VTA every 10 s for 10 trials at each frequency. To validate spike collision, we varied the collision interval (0, 1, 2, 5, 10 and 20 ms) between orthodromic and antidromic photostimulation. Each collision interval was repeated every 5 s for a total of 10 trials. Data acquisition and analysis was performed using pCLAMP software (Molecular Devices). Placements of recording electrode tips within the vBNST and optical fibres within the VTA were verified with histological examination of brain tissue after the experiments.

Patch-clamp electrophysiology. Brain slices preparation and general methods for patch-clamp electrophysiology were conducted as previously described^{10,13}, with the following changes. To examine both vBNST postsynaptic glutamatergic and GABAergic currents, C57BL/6J mice were injected with AAV5 CaMKII α ::Chr2-eYFP to nonspecifically target vBNST–VTA projection neurons (CaMKII α ^{vBNST–VTA}::Chr2). For whole-cell voltage recordings (EPSCs and IPSCs) from VTA neurons, electrodes (2–4 M Ω electrode resistance) contained (in mM): 117 caesium methanesulphonate, 20 HEPES, 0.4 EGTA, 2.8 NaCl, 5 TEA, 2 Mg-ATP, 0.2 Na-GTP, pH 7.2–7.4, 275–285 mOsm. The caesium methanesulphonate internal solution also contained the selective NMDA (*N*-methyl-D-aspartate) antagonist MK-801 (1.125 mM). VTA neurons were clamped at holding potential (V_h) of -70 mV (reversal potential (E_{rev}) for GABA_A receptors) and $+10$ mV (E_{rev} for AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors) to examine both glutamatergic and GABAergic postsynaptic currents, respectively, within the same neuron. Photostimulation (5-ms pulses of 1–2 mW, 473-nm light delivery via LED through a $\times 40$ microscope objective) was used at both voltages. In a subset of neurons ($n = 4$ of 11 neurons), tetrodotoxin (1 μ M) and 4-aminopyridine (1 mM) were bath-applied to isolate mono-synaptic currents.

To isolate vBNST glutamatergic and GABAergic postsynaptic currents, *Vglut2*^{vBNST–VTA}::Chr2 and *Vgat*^{vBNST–VTA}::Chr2 mice were used for recordings. For whole-cell voltage-clamp recordings of VTA EPSCs from *Vglut2*^{vBNST–VTA}::Chr2 mice, once stable light-evoked EPSCs were achieved, 10 μ M DNQX was bath-applied. For whole-cell voltage-clamp recordings of VTA IPSCs from *Vgat*^{vBNST–VTA}::Chr2 mice, once stable light-evoked IPSCs were achieved, 10 μ M gabazine was bath-applied.

Ex vivo validation of photoinhibition of VTA GABA neurons. For current clamp recordings to show hyperpolarization of membrane voltage from VTA GABA neurons after eNpHR3.0 inhibition, *Vgat-ires-cre* mice were transduced with Cre-inducible eNpHR3.0 under the control of the EF1 α promoter in the VTA. Electrodes (2–4 M Ω) contained (in mM): 130 K-gluconate, 10 KCl, 10 HEPES, 10 EGTA, 2 MgCl₂, 2 Mg-ATP, 0.2 Na-GTP, pH 7.2–7.4, 275–285 mOsm. VTA neurons were maintained at ~ 60 mV. For photoinhibition, 500-ms pulses of 5–8 mW, 532-nm light delivery via a solid-state laser coupled to an optical fibre positioned in the brain slice.

Unpredictable foot-shock model. Mice were placed in sound-attenuated mouse behavioural chambers (Med Associates), where an unpredictable foot shock was used as an aversive stimulus. A house light and white noise signalled the start of the trial and remained on throughout the entire 20-min foot-shock session (contextual cues). Each unpredictable foot shock was 0.75 mA in intensity and 500 ms in duration on a variable interval (VI60) schedule. Mice received approximately 20 unpredictable foot shocks during the entire 20-min session.

For the *in vivo* electrophysiology experiments, *Vglut2*^{vBNST–VTA}::Chr2 and *Vgat*^{vBNST–VTA}::Chr2 mice implanted with a multielectrode array in the vBNST and an optical fibre in the VTA were placed in the unpredictable foot-shock context in which they received the visual and auditory contextual cues in the absence of foot shock for 20 min (contextual cue exposure before shock association). Five days after the unpaired contextual cue session, mice were run in the standard unpredictable foot-shock session (approximately 20 unpredictable foot shocks, 0.75 mA, 500 ms) for 5 consecutive days as stated above, which included paired presentation of the contextual cues. Seven days after the fifth unpredictable foot-shock session, mice were placed back into the unpredictable foot-shock model, in which they received the visual and auditory contextual cues (house light and white noise) in the absence of foot shock for 20 min.

***In vivo* electrophysiology.** Neural activity was recorded using an Omniplex recording system (Plexon Instruments). Signals from each electrode in the array were referenced to ground, and recordings were performed in differential mode to subtract artefacts unrelated to neural activity. Acquired data was band-pass filtered between 0.1 and 8,000 Hz. Spike sorting was performed offline using Offline sorter (Plexon Instruments), as previously described^{31,32}. In brief, discrimination of individual units was performed offline using principal component analysis to separate

individual units from the same electrode. In addition, auto- and cross-correlograms, firing characteristics, and inter-spike interval distributions were examined to ensure units were well isolated. In addition, timestamp data to signify the start and end of foot-shock sessions, and the delivery of light pulses to optical fibres were synchronized with electrophysiological data. Sorted waveforms were further processed in NeuroExplorer (Nex Technologies) to extract unit timestamps and relevant events. NeuroExplorer-extracted timestamps were exported to MATLAB for further data processing and statistical testing. Neuronal units were included in the data if the signal-to-noise ratio was high, and the mean firing rate was between 0.5 and 25 Hz during baseline recording periods. During the time epoch surrounding the individual delivered foot shocks, large shock artefacts were readily apparent. Although these were easily isolated and excluded from the analysed waveform data using offline spike sorting, neuronal firing responses in the 500–1,000 ms after foot-shock onset could not be reliably quantified. Thus, data collected during these time epochs were excluded from analysis. This represents an approximate loss of 0.8–1.66% of the collected data during the foot-shock session. Resulting data was binned in 30-s epochs to minimize the skewing of our results due to data lost by the shock artefacts. In addition, recording sites was verified histologically using electrolytic lesions at 200 μ A for 5 s.

To identify units originating from vBNST projection neurons, 5-ms light pulses were delivered to the VTA to antidromically stimulate vBNST projection neurons that innervated the area. Light pulses were delivered in 10-s intervals for 20 trials starting 40 min after the end of the unpredictable foot-shock session. Recorded vBNST units were classified as light-responsive, and thus VTA-projecting, if they met both of the following two criteria: (1) the latency of the first spike after light stimulation onset was less than 20 ms for $\geq 20\%$ of the trials; and (2) light-evoked and spontaneous waveform shapes had a correlation coefficient of >0.90 . To compare light-evoked and spontaneous waveforms from units, light-evoked waveform characteristics were defined using the average waveform shape and average principal component values (PC1–3) of the first spike after photostimulation onset from each successful trial where a waveform was collected within 20 ms after light onset. This subset of light-evoked waveforms was then averaged together for a given unit, and compared to a subset of spontaneous, non-light-evoked waveforms that occurred immediately before the onset of light stimulations (pre-stimulation waveforms) and the first collected waveforms occurring after the 20-ms interval after the offset of photostimulations (post-stimulation waveforms). The correlation between each average waveform shape over the three time epochs (pre-stimulation, during stimulation and post-stimulation) was then calculated using Pearson's product-moment coefficient as well as their average principle component values.

VTA-projecting vBNST neuronal units were then further classified, dependent on their firing response to the foot-shock session into three categories: foot-shock session-excited, foot-shock session-inhibited, and foot-shock session-no effect. To clarify, foot-shock session-excited, foot-shock session-inhibited, and foot-shock session-no effect refers to the activity of a neural unit in response to the collective aversive experience, not to individual foot shocks. To assess the firing rate of a particular neuron, each spike from 10 min before the foot-shock session to 20 min after the end of the foot-shock session was binned into 30-s bins. Firing rate was then normalized to the mean firing rate during the 10-min before the start of the 20-min foot-shock session using z-scores. Neurons were classified as foot-shock-session-excited if their average z-score during the 20-min foot-shock session was greater than 1. Likewise, neurons were classified as foot-shock-session-inhibited if their average z-score during the shock session fell below -1 . All other units in which the z-score did not exceed an absolute value of 1 during the 20-min foot-shock session were classified as no effect. Neural activity was recorded from the same mice during the cue exposure, foot shock, and cue re-exposure sessions, therefore we were able to record activity reliably from the same population of neurons during each of the three sessions.

Photostimulation and photoinhibition during the RTPP test. *Vglut2*^{vBNST-VTA}::ChR2, *Vgat*^{vBNST-VTA}::ChR2, *Vgat*^{VTA}::eNpHR3.0 and littermate controls were implanted with optical fibres above the VTA and were run in the RTPP test. See ref. 10 for further details.

Intra-VTA injection of antagonists and photostimulation during the RTPP test. A separate cohort of *Vglut2*^{vBNST-VTA}::ChR2 and *Vgat*^{vBNST-VTA}::ChR2 mice was unilaterally implanted with a 26-gauge cannula coupled to an optical fibre aimed above the VTA. All mice were placed in a custom-made place preference arena and were run in the RTPP test to achieve a baseline measurement. Two days after the baseline session, *Vglut2*^{vBNST-VTA}::ChR2 mice were injected with either 0.3 μ l of vehicle (saline) or a cocktail of selective glutamate antagonists (0.1 μ g AP5, 0.001 μ g DNQX in saline), and *Vgat*^{vBNST-VTA}::ChR2 mice were injected with either 0.3 μ l of vehicle (saline) or a selective GABA_A antagonist (0.001 μ g gabazine) into the VTA in a counter-balanced design (all drugs from Tocris). The injector needle (33-gauge steel tube, McMaster-Carr) extended approximately

1 mm past the cannula to ensure drug delivery 0.5 mm below the optical fibre. All mice were infused at a rate of 0.1 μ l per minute. The injector remained in place for approximately 2 min after infusion to ensure proper diffusion of drug into the VTA. Immediately after the microinjection procedure, all mice were placed into the RTPP chamber. Mice had 2 days off in between each VTA microinjection.

Photostimulation of projections during open-field testing. *Vglut2*^{vBNST-VTA}::ChR2 and *Vglut2*^{vBNST-VTA}::control mice were examined in a custom-made open-field arena (25 \times 25 \times 25-cm white plexiglass arena) for 35 min. After a baseline period of 5 min, all mice received constant 20-Hz photostimulation. Immediately after the 20-min photostimulation epoch, all mice had a 10-min period in which they received no photostimulation. The centre zone was defined as the centre 156 cm² (25% of the entire arena). Corner zones were defined as the 39 cm² in each corner. The 35-min session was recorded with a CCD camera that was interfaced with Ethovision software (Noldus Information Technologies). Time spent in the corner and the centre of the open-field apparatus was recorded. Heat maps and post-acquisition processing were conducted in MATLAB (Mathworks).

Photostimulation of projections during sucrose self-administration. *Vglut2*^{vBNST-VTA}::ChR2 and *Vglut2*^{vBNST-VTA}::control mice with optical fibres implanted above the VTA were first food restricted to 90% of their free-feeding weight. They were then placed in standard mouse operant chambers to nose poke for a 15% (w/v) sucrose solution on a fixed-ratio 1 schedule in a 30-min session. Once stable nose-poking behaviour for 15% sucrose was observed (approximately 100 active nose pokes on at least two consecutive days), all mice received constant 20-Hz photostimulation during the entire 30-min sucrose session.

Optical self-stimulation of *Vgat*^{vBNST-VTA}::ChR2 projections. *Vgat*^{vBNST-VTA}::ChR2 and *Vgat*^{vBNST-VTA}::control mice with optical fibres implanted above the VTA were trained in one 30-min session to nose poke on a fixed-ratio 1 schedule for optical self-stimulation of the *Vgat*^{vBNST-VTA}::ChR2 projections in standard mouse operant chambers (Med Associates). Each nose poke resulted in a single 3-s 20-Hz optical pulse train. After the 1-day 20-Hz training session, mice were run daily at each photostimulation frequency (1, 5, 10, 20 and 40 Hz) in a counter-balanced design.

Optical self-inhibition of *Vgat*^{VTA}::eNpHR3.0 neurons. *Vgat*^{VTA}::eNpHR3.0 and *Vgat*^{VTA}::control mice with optical fibres implanted above the VTA were trained in one 30-min session to nose poke on a fixed-ratio 1 schedule for photo-inhibition of VTA GABAergic cell bodies in standard mouse operant chambers as described above (Med Associates).

Photostimulation and photoinhibition during the EPM test. *Vgat*^{vBNST-VTA}::ChR2, *Vgat*^{VTA}::eNpHR3.0, *Vgat*^{VTA}::control, and *Vgat*^{vBNST-VTA}::control mice were run in the EPM test to assay anxiety-like behaviour. Activity and location was recorded for 5 min (baseline). After this 5-min baseline period, *Vgat*^{vBNST-VTA}::ChR2 and *Vgat*^{vBNST-VTA}::control mice received constant 20-Hz photostimulation for 5 min, whereas *Vgat*^{VTA}::eNpHR3.0 and *Vgat*^{VTA}::control mice received constant inhibition for 5 min. Immediately after the 5-min photostimulation or photoinhibition epoch, all mice had a 5-min period in which they received no light delivery.

Photostimulation during foot-shock, freezing and anxiety-like behaviour. *Vgat*^{vBNST-VTA}::ChR2 and *Vgat*^{vBNST-VTA}::control mice with optical fibres implanted above the VTA were run in a modified foot-shock test as described earlier. In brief, mice were placed into sound-attenuated mouse chambers (Med Associates) for a 5-min baseline period. After the 5-min baseline period, a house light and white noise were activated and mice received the same foot-shock protocol as described above. Furthermore, during the 20-min shock session, all mice received constant 20-Hz photostimulation. A separate cohort of mice (*Vgat*^{vBNST-VTA}::ChR2 and *Vgat*^{vBNST-VTA}::control) received constant 20-Hz photostimulation of this pathway in the absence of foot shock. Immediately after the 20-min foot-shock and photostimulation epoch, all mice had a 5-min period in which they received no foot shock or photostimulation while still exposed to contextual cues, to assay freezing behaviour. Freezing was defined as the total lack of any movement, except respiration, for a period of 2 s. The 30-min test session was recorded with a CCD camera that was interfaced with Ethovision software (Noldus Information Technologies). The time frozen during the 5-min period immediately after the foot shock and photostimulation session was recorded. Approximately 3 h after the foot-shock and photostimulation session or just the photostimulation session in the absence of foot shock, mice were run on the EPM test to assay anxiety-like behaviour for 5 min.

- Roitman, M. F., Wheeler, R. A. & Carelli, R. M. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron* **45**, 587–597 (2005).
- Tye, K. M. & Janak, P. H. Amygdala neurons differentially encode motivation and reinforcement. *J. Neurosci.* **27**, 3937–3945 (2007).

Diverse and heritable lineage imprinting of early haematopoietic progenitors

Shalin H. Naik^{1†}, Leila Perié^{1,2}, Erwin Swart¹, Carmen Gerlach¹, Nienke van Rooij¹, Rob J. de Boer² & Ton N. Schumacher¹

Haematopoietic stem cells (HSCs) and their subsequent progenitors produce blood cells, but the precise nature and kinetics of this production is a contentious issue. In one model, lymphoid and myeloid production branch after the lymphoid-primed multipotent progenitor (LMPP)¹, with both branches subsequently producing dendritic cells². However, this model is based mainly on *in vitro* clonal assays and population-based tracking *in vivo*, which could miss *in vivo* single-cell complexity^{3–7}. Here we avoid these issues by using a new quantitative version of ‘cellular barcoding’^{8–10} to trace the *in vivo* fate of hundreds of LMPPs and HSCs at the single-cell level. These data demonstrate that LMPPs are highly heterogeneous in the cell types that they produce, separating into combinations of lymphoid-, myeloid- and dendritic-cell-biased producers. Conversely, although we observe a known lineage bias of some HSCs^{11–14}, most cellular output is derived from a small number of HSCs that each generates all cell types. Crucially, *in vivo* analysis of the output of sibling cells derived from single LMPPs shows that they often share a similar fate, suggesting that the fate of these progenitors was imprinted. Furthermore, as this imprinting is also observed for dendritic-cell-biased LMPPs, dendritic cells may be considered a distinct lineage on the basis of separate ancestry. These data suggest a ‘graded commitment’ model of haematopoiesis, in which heritable and diverse lineage imprinting occurs earlier than previously thought.

We sought to determine the ability of individual LMPPs to produce an array of different cell types, including lymphoid and myeloid cells, but also three major dendritic cell subsets (Supplementary Table 1). To examine this, LMPPs were sorted similarly to their original description¹ (Supplementary Fig. 2a), and labelled with a library of heritable DNA barcodes, so that transduced cells typically received only one unique barcode (Fig. 1a, Supplementary Information and Supplementary Fig. 1c). Labelled cells were transferred into partially irradiated mice, and were allowed to proliferate and differentiate. Subsequently, the progeny cell types of interest were isolated and their barcode repertoire assessed by a new multiplexed next-generation sequencing strategy, enabling a sensitive and semi-quantitative readout of lineage relationships (Supplementary Figs 1c and 3, and Supplementary Information). Barcodes that were present in only one cell type revealed the output of mono-outcome progenitors, whereas those found in two or more cell types reflected the output of bi- or multi-outcome progenitors. This modification of our original strategy^{8–10} is similar to recent methods^{15,16}, and represents an addition to pioneering retroviral tagging¹² and single-cell transplantation studies^{11,17,18}.

As the derivation of myeloid and lymphoid cells from LMPPs kinetically overlaps at day 14 after LMPP transfer¹⁹, this time was chosen to collect cell types. All of these cell types, except T cells, contained green-fluorescent-protein-expressing (GFP⁺) donor progeny (Supplementary Fig. 2b and Supplementary Information). Barcode signatures were assessed for every cell type and visualized on a clustered heatmap (Fig. 1b).

Engraftment of the barcoded LMPPs was $71 \pm 8.3\%$ (number of barcodes detected divided by number of barcoded LMPPs injected,

\pm s.e.m., $n = 7$). However, the bulk of each cell type (90%) was produced by very different numbers of LMPPs, providing the first evidence for heterogeneous LMPP behaviour (Supplementary Fig. 4a and Supplementary Table 2). To examine further this disparity, we asked

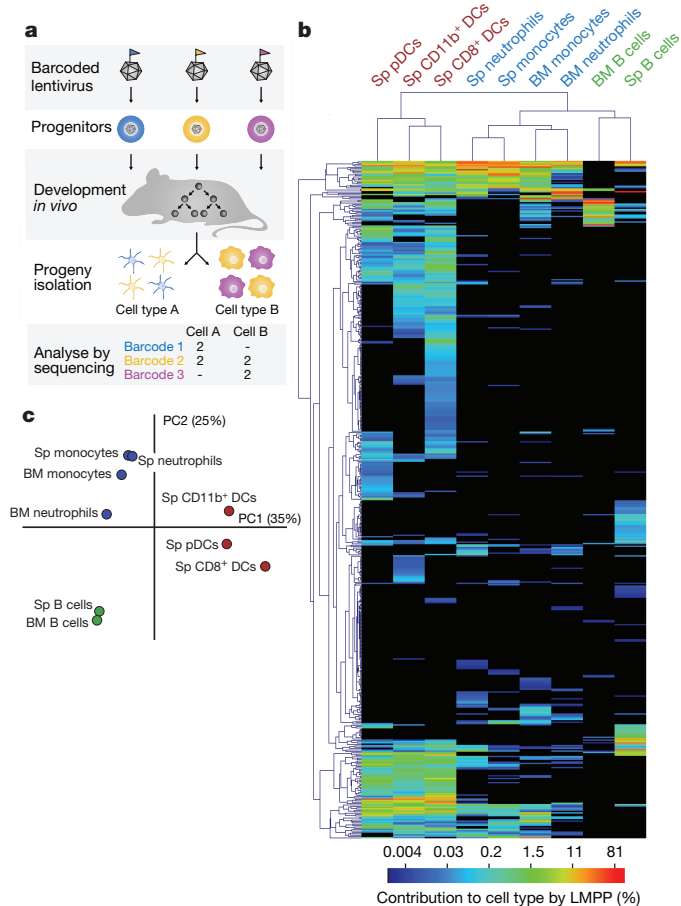


Figure 1 | Single LMPP output to different cell types. **a**, Cellular barcoding protocol. Progenitors are transduced so that each receives a unique and heritable DNA barcode. After differentiation, resulting cell types are isolated and analysed for their barcode repertoire. **b**, Heatmap representation of the output of individual LMPPs (rows) to different cell types (columns) after 2 weeks (clustering using complete linkage and Euclidian distance). Each element represents the relative contribution of an LMPP to each cell type (see scale). Cell types are colour-coded by the first three branches in horizontal clustering. **c**, Principal component analysis of cell types with percentage variation explained per coordinate in brackets. Two pooled mice from experiment 3 that received 300 barcode-labelled cells each. B, B cells; BM, bone marrow; DCs, dendritic cells; PC1, principal component 1; pDC, plasmacytoid dendritic cell; Sp, spleen.

¹Division of Immunology, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands. ²Theoretical Biology and Bioinformatics, Utrecht University, 3584 CH Utrecht, The Netherlands.

[†]Present address: Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia.

how many output cell types each LMPP produced, and found that a marked proportion (75%) of LMPPs contributed to only 1, 2 or 3 of the cell types tested (Supplementary Fig. 4b). This was not a result of poor recovery of barcodes, as shown by technical replicates (Supplementary Fig. 3a) and by the finding that dendritic cells, which have the lowest cell numbers, had the largest number of different barcodes.

We subsequently applied both clustering and principal component analysis (PCA) to look for patterns in the data. These analyses separated the output cell types into three major groups; the dendritic cell subtypes, the myeloid cell types and B cells (horizontal clustering in Fig. 1b; see also Fig. 1c). This grouping was confirmed by bootstrapping and was reproducible across experimental repeats (Supplementary Figs 5 and 6).

As clustering of LMPPs did not give rise to well-defined groups (Fig. 1b, vertical clustering), we used PCA to order the progenitors. This segregated LMPPs into three main groups, with lower densities of 'intermediate' progenitors between them (Supplementary Fig. 7a and Supplementary Video 1). Colour coding of these progenitors by their three-dimensional principal component coordinates revealed that the different groups corresponded to producers of dendritic cells, B cells or myeloid cells, or combinations of these (Supplementary Fig. 7b).

On the basis of the observation that LMPPs showed distinct patterns of cellular output, we generated a hand-tailored classifier to assign each LMPP to one of seven possible classes of lineage bias (Fig. 2a and Supplementary Fig. 8). Almost one-half of the LMPPs were classified as dendritic-cell biased, 10% of the progenitors contributed primarily to B cells, 10% primarily to myeloid cells, and just 3% of the LMPPs were multi-outcome progenitors (Fig. 2a). Notably, lineage-biased LMPPs were not only numerous but also contributed to a substantial fraction of total dendritic-cell, B-cell and myeloid-cell progeny (Fig. 2b, left panel, and Supplementary Figs 1a, b and 9). In contrast, the collective cellular output of the few multi-outcome LMPPs was small.

The pronounced heterogeneity in LMPP output prompted us to examine whether such patterns are observed at an earlier stage. We therefore barcode-labelled long-term-reconstituting HSCs²⁰ (Supplementary Fig. 2a) and determined their output after 16 weeks.

Only a small fraction of transplanted HSCs was a major contributor to cell output (60–80%), and these HSCs were mainly multi-outcome (Fig. 2b, right, and Supplementary Fig. 10). Although some HSCs had a lymphoid- or myeloid- plus dendritic-cell bias, the combined cellular output of these lineage-biased HSCs was small, consistent with previous studies^{11–13,15,16}. Thus, although some lineage bias exists amongst HSCs, major diversification in output patterns only manifests subsequently.

An explanation for the observed LMPP heterogeneity could be a difference in the kinetics of lineage production. For example, dendritic cells and myeloid cells could derive from true multi-outcome LMPPs, but with a transient wave of myeloid output preceding their output to dendritic cells. However, although some increase in myeloid-cell-biased LMPPs was observed when assessed at day 10 (the time of their peak output from LMPPs¹⁹), a large number of dendritic-cell-biased progenitors was still detected at this time point, and no increase in joint myeloid- plus dendritic-cell-biased LMPPs was observed relative to day 14 (Fig. 3a). Therefore, many dendritic cells were derived from LMPPs that did not produce myeloid cells at day 10.

A related concern was that the LMPPs that were shown to produce dendritic cells and/or myeloid cells at day 14 could produce a subsequent cohort of B cells at a time point at which dendritic-cell and myeloid-cell output has largely disappeared (day 28). To address this concern we carried out partial splenectomies to assess the kinship between 'late' B cells and 'early' dendritic cells, B cells and myeloid cells (partial splenectomy lineage-tracing controls in Supplementary Fig. 11a). This analysis showed that B cells at both time points were sufficiently related to each other, and unrelated to myeloid cells and dendritic cells, to form a distinct cluster (Fig. 3b and Supplementary Fig. 11b, c). Furthermore, only a small number of LMPPs produced myeloid cells and dendritic cells at day 14, and B cells at day 28, and this had little effect on the distribution of LMPP classes when data from both time points were combined (Fig. 3a). Thus, the observed heterogeneity in LMPP output is not explained by the kinetics of cell production.

The above data demonstrate that most LMPPs have a restricted lineage output. However, this does not address whether this fate is imprinted, or determined stochastically or by the microenvironment after cell transfer (Supplementary Fig. 12). In particular, the use of irradiated recipient mice, necessary for sufficient clonal expansion for reproducible barcode analysis (data not shown), could potentially override an intrinsic multi-lineage potential.

We reasoned that comparison of the output patterns of siblings derived from single LMPPs should enable us to test this. In particular, if the lineage bias of siblings were the same after transfer into two separate mice ('conserved' fate), this would be indicative of pre-existing imprinting before their *in vitro* clonal expansion. To address this, we barcode-labelled LMPPs, pre-expanded them as a population *in vitro* and, after splitting, transferred them into pairs of recipient mice (Supplementary Fig. 13a). Following *in vivo* differentiation, barcode repertoires were assessed in dendritic cells, myeloid cells and B cells, focusing on those barcodes that were shared between two mice of a pair, as only these allow conservation of fate to be tested (Fig. 4a and Supplementary Fig. 13b, c). To remove the effect of differential engraftment or expansion, output per progenitor in each recipient was normalized and then analysed by clustering (Fig. 4b).

Importantly, in all analyses the same cell types from two mice of a pair clustered, rather than cell types from each individual mouse, indicating that the intrinsic lineage preference of siblings was dominant. Indeed, conserved fate was greater than expected by chance for three of the four mouse pairs (Fig. 4c, pair two had poor engraftment) and we found similar proportions of conserved bias in the B-, dendritic-cell, B- plus dendritic-cell, and myeloid- plus dendritic-cell lineages (B, DC, B/DC and M/DC; Fig. 4d and Supplementary Fig. 13d). Notably, whereas many siblings only produced dendritic cells in pairs of mice, and others produced both myeloid cells plus dendritic cells in mouse pairs, siblings

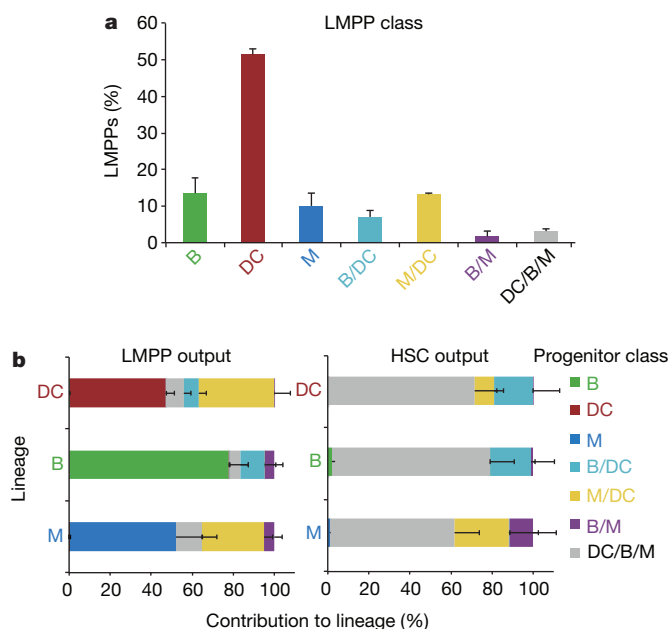


Figure 2 | Classification of lineage bias in LMPPs and HSCs. **a**, Proportion of LMPPs classified by the indicated lineage bias (average \pm s.e.m., $n = 7$). **b**, Quantitative contribution of these classes to each lineage for LMPPs (left, average \pm s.e.m., $n = 7$) and HSCs (right, average \pm s.e.m., $n = 4$, two independent experiments for each). Note the difference in contribution by multi-outcome cells that produce dendritic, B and myeloid cells (DC/B/M, grey) between HSCs and LMPPs.

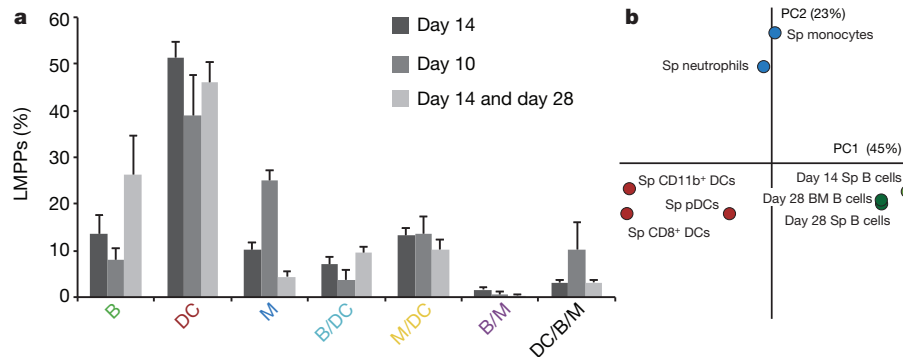


Figure 3 | Kinetics of lineage production do not explain LMPP fate heterogeneity. **a**, Classification of barcoded LMPPs that give rise to lineages collected after 10 or 14 days, or from combined data of cell types isolated at days 14 and 28 by partial splenectomy. Data are pooled from three to seven

mice from two separate experiments each (average \pm s.e.m.). **b**, PCA of cell types derived from organs obtained by partial splenectomy. One representative experiment.

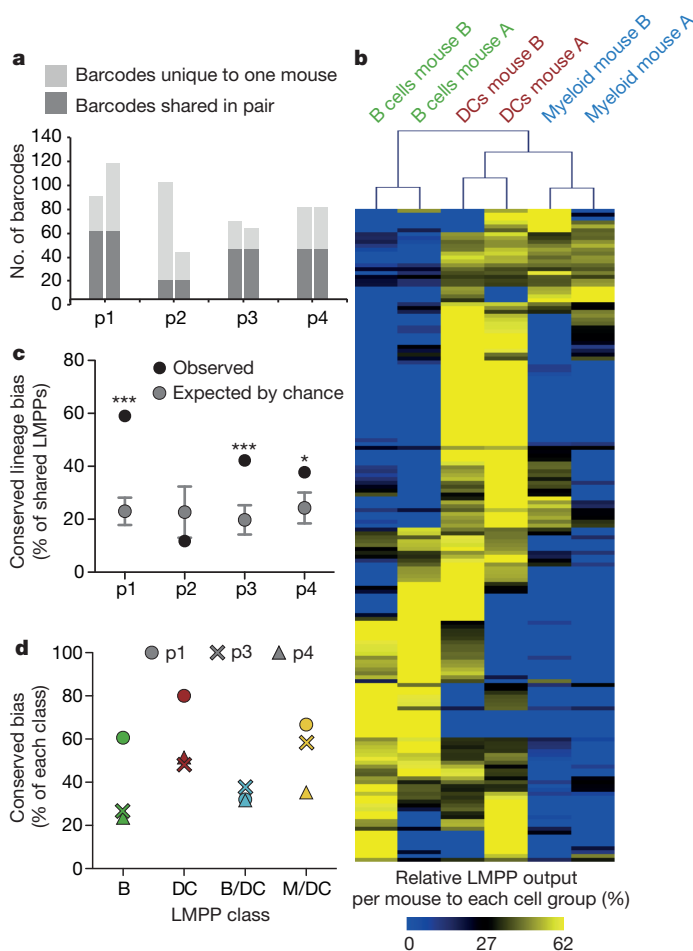


Figure 4 | Pre-existing and heritable lineage imprinting within LMPPs. Barcode-labelled LMPPs were pre-expanded *in vitro*, split, injected into mouse pairs (p), and barcode repertoires were determined after 2 weeks. **a**, Barcodes in each mouse of a pair that were either unique, or present in both (shared). **b**, Clustered heatmap (Pearson correlation) of the output of shared barcodes, normalized per progenitor for each mouse (pooled data from pairs 1–4). **c**, Experimentally observed percentage of shared LMPPs demonstrating the same lineage bias ('conserved' bias) versus the percentage expected by chance (average \pm s.d. of 1,000 randomizations). **d**, LMPP siblings with conserved bias (percentage of each LMPP class). Pair 2 were excluded owing to poor engraftment. Data from pairs 1, 3 and 4 fall outside of the *95% and ***99.9% confidence intervals of the randomization distribution.

that produced dendritic cells in one mouse, and myeloid plus dendritic cells in the other were not more common than expected by chance (Supplementary Fig. 13d). Thus, although we cannot exclude the possibility that an irradiated niche may increase clonal expansion of progenitors, these data directly demonstrate that the heterogeneity in output patterns of LMPPs, including towards the dendritic-cell lineage, is at least partially imprinted. Those remaining siblings that produced non-conserved output between two mice may represent LMPPs without pre-determined fate or for which niche effects did have a role.

Our findings, which suggest that diverse lineage imprinting occurs between the HSC and LMPP stages, lend support to previous findings^{1,3,5–7,11,12,14,18,21,22}. On the basis of these data we argue in favour of a revised model of haematopoiesis in which heritable lineage imprinting is initiated as early as at the HSC level and progressively increases throughout haematopoiesis, ultimately giving rise to the differentiated cell types. Thus, unlike the classic bifurcation model, haematopoiesis seems to occur through a process of 'graded commitment'^{23,24}, a model with similar elements to others that have been proposed previously^{7,13,25,26}. The observation of diverse imprinting patterns in LMPPs, including towards what we now consider a separate 'dendritic cell' lineage^{24,27}, makes the identification of the nature, expression and timing of 'early-acting' factors controlling this behaviour an important next challenge. Correlation of the *in vivo* lineage bias and transcriptional profiles of siblings derived from the same progenitor should enable this issue to be addressed.

METHODS SUMMARY

Mice, progenitor isolation and barcode labelling. C57BL/6J (CD45.2) donor mice, and C57BL/6 Pep3b (CD45.1) recipient mice irradiated with 5 gray were used in all experiments. LMPPs and HSCs were sorted using similar methods to those described previously^{1,20} (Supplementary Fig. 3). The barcode library⁸ was cloned to produce lentivirus for transduction. Progenitors were transduced for 6 h with 50 ng ml⁻¹ stem cell factor (SCF) and injected intravenously into recipient mice or, for sibling analyses, cultured for a further 4 days with additional 30 ng ml⁻¹ Flt3 ligand, before the cultures were split in two and introduced through intrasplenic injection into pairs of mice.

Cell isolation. Cell suspensions were derived from organs of killed mice, or from live mice after partial splenectomy. After sequential magnetic-bead enrichment for dendritic cells, myeloid cells or B cells, and subsequent staining, the indicated cell types were isolated by flow cytometry (Supplementary Fig. 2). For analyses of these experiments, and the analyses at day 10, CD45.2⁺-enriched total dendritic cells, myeloid cells or B cells were sorted.

Polymerase chain reaction and sequencing. Sorted cell samples were lysed in Viagen buffer and split into technical replicates⁹. Barcodes were amplified by polymerase chain reaction (PCR)⁸, and then underwent a second PCR that tagged each sample with a unique index primer (Supplementary Fig. 1c). Tagged PCR products were pooled and sequenced by next-generation sequencing.

Bioinformatics. Raw data were first filtered to ensure that only samples of sufficient quality were assessed further (Supplementary Information). Resulting data were collated in a table of normalized barcode read counts in each cell type, and this

formed the basis for further bioinformatics analysis (Supplementary Information). To classify individual progenitors by their lineage bias, we applied an additional normalization per progenitor in each mouse, thereby enabling categorization into LMPPs with mono-, bi- and tri-potency for the B-cell, dendritic-cell and myeloid lineages. Potency was assigned if more than 2.5% of the output of an LMPP was towards that lineage.

Full Methods and any associated references are available in the online version of the paper.

Received 2 April 2012; accepted 14 February 2013.

Published online 3 April 2013.

- Adolfsson, J. *et al.* Identification of Flt3⁺ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. *Cell* **121**, 295–306 (2005).
- Shortman, K. & Naik, S. H. Steady-state and inflammatory dendritic cell development. *Nature Rev. Immunol.* **7**, 19–30 (2007).
- Arinobu, Y. *et al.* Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. *Cell Stem Cell* **1**, 416–427 (2007).
- Igarashi, H., Gregory, S. C., Yokota, T., Sakaguchi, N. & Kincade, P. W. Transcription from the RAG1 locus marks the earliest lymphocyte progenitors in bone marrow. *Immunity* **17**, 117–130 (2002).
- Akashi, K. Lineage promiscuity and plasticity in hematopoietic development. *Ann. NY Acad. Sci.* **1044**, 125–131 (2005).
- Naik, S. H. *et al.* Development of plasmacytoid and conventional dendritic cell subtypes from single precursor cells derived *in vitro* and *in vivo*. *Nature Immunol.* **8**, 1217–1226 (2007).
- Mansson, R. *et al.* Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26**, 407–419 (2007).
- Schepers, K. *et al.* Dissecting T cell lineage relationships by cellular barcoding. *J. Exp. Med.* **205**, 2309–2318 (2008).
- Schumacher, T. N., Gerlach, C. & van Heijst, J. W. Mapping the life histories of T cells. *Nature Rev. Immunol.* **10**, 621–631 (2010).
- van Heijst, J. W. *et al.* Recruitment of antigen-specific CD8⁺ T cells in response to infection is markedly efficient. *Science* **325**, 1265–1269 (2009).
- Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs *in vivo*. *Cell Stem Cell* **1**, 218–229 (2007).
- Lemischka, I. R. What we have learned from retroviral marking of hematopoietic stem cells. *Curr. Top. Microbiol. Immunol.* **177**, 59–71 (1992).
- Muller-Sieburg, C. E. & Sieburg, H. B. The GOD of hematopoietic stem cells: a clonal diversity model of the stem cell compartment. *Cell Cycle* **5**, 394–398 (2006).
- Challen, G. A., Boles, N. C., Chambers, S. M. & Goodell, M. A. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF- β 1. *Cell Stem Cell* **6**, 265–278 (2010).
- Gerrits, A. *et al.* Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618 (2010).
- Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells *in vivo* using high-throughput sequencing in conjunction with viral genetic barcoding. *Nature Biotechnol.* **29**, 928–933 (2011).
- Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
- Sieburg, H. B. *et al.* The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets. *Blood* **107**, 2311–2316 (2006).
- Forsberg, E. C., Serwold, T., Kogan, S., Weissman, I. L. & Passegue, E. New evidence supporting megakaryocyte–erythrocyte potential of flk2/flt3⁺ multipotent hematopoietic progenitors. *Cell* **126**, 415–426 (2006).
- Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Hu, M. *et al.* Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11**, 774–785 (1997).
- Miyamoto, T. *et al.* Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev. Cell* **3**, 137–147 (2002).
- Naik, S. H. Demystifying the development of dendritic cell subtypes, a little. *Immunol. Cell Biol.* **86**, 439–452 (2008).
- Naik, S. H. *Dendritic Cell Development, Lineage Issues and Haematopoiesis at the Single Cell Level*. 79–120 (Nova Press, 2009).
- Welner, R. S., Pelayo, R. & Kincade, P. W. Evolving views on the genealogy of B cells. *Nature Rev. Immunol.* **8**, 95–106 (2008).
- Ceredig, R., Rolink, A. G. & Brown, G. Models of haematopoiesis: seeing the wood for the trees. *Nature Rev. Immunol.* **9**, 293–300 (2009).
- Robbins, S. H. *et al.* Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol.* **9**, R17 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Filion, U. Braunschweig, L. Pagie, M. Hauptmann, P. Lio, R. van der Wath and the NKI Genomics Facility for computational assistance; J. van Heijst, J. Rohr and J. Urbanus for useful discussions; <http://www.josharris.com> for illustrations in Fig. 1a and Supplementary Fig. 1; NKI cytometry and animal facilities; D. Leone, Z. Ping and M. Lodder for technical assistance; and P. Hodgkin, K. Duffy and J. Coquet for critical reading of the manuscript. This work was supported by ERC grant LIFE-HIS-T and HFSP grant RGP0060/2012. S.H.N. was supported by the National Health and Medical Research Council Australia, Marie Curie Incoming International FP6 Fellowship, and the Leukemia and Lymphoma Society; and L.P. was supported by a Marie Curie Intra European FP7 Fellowship and the Bettencourt Schueller Foundation.

Author Contributions S.H.N. conceived, developed, executed and analysed experiments; L.P. developed and carried out data analysis, advised by R.J.d.B.; E.S., C.G. and N.v.R. contributed to lineage-tracing technology development. T.N.S. conceived the approach and advised on data analysis and interpretation. All authors discussed results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.H.N. (naik.s@wehi.edu.au) or T.N.S. (t.schumacher@nki.nl).

METHODS

Mice. C57BL/6J (CD45.2) mice were used as progenitor donors, and C57BL/6 Pep3b (CD45.1) mice were used as recipients. In all experiments, mice that were used were 7–16 weeks old. Mice were supplied by Charles River Laboratories and housed at the Netherlands Cancer Institute (NKI). Mice were killed by cervical dislocation before organs were collected. Recipient mice were irradiated with a single 5-Gy dose. All animal experiments were carried out in accordance with national guidelines and were approved by the Experimental Animal Committee of the NKI.

Progenitor isolation and barcode labelling. Tibia, femur and ilium were isolated from mice that had been killed, and were flushed with RPMI 1640 medium (Gibco) with 10% fetal calf serum (FCS) to a single-cell suspension. Haematopoietic progenitors were isolated by staining with anti-CD117 allophycocyanin (APC), followed by magnetic-bead enrichment using anti-APC beads. Cells were stained with additional antibodies against Sca1, CD135, CD16/32, CD127 and, when sorting for HSCs, CD150. LMPPs were sorted as CD16/32⁺CD127⁺CD117^{hi}Sca1⁺CD135^{hi} and HSCs as CD16/32⁺CD127⁺CD117^{hi}Sca1⁺CD135^{lo}CD150⁺, similar to the original descriptions^{1,20}. The barcode library⁸ was cloned from the retroviral pMX vector into a lentiviral pLL vector²⁸ for lentivirus generation, concentrated 100-fold using 300,000 MWCO Amicon columns (Millipore) and stored in small aliquots at –80 °C. LMPPs were resuspended at no more than 1×10^5 in 100 μ L StemSpan (StemCell Technologies), supplemented with 50 ng ml^{–1} SCF (Peprotech), and 100 μ L were aliquoted per well in a round-bottomed well of a 96-well plate. A small volume of virus, predetermined to give 5–10% transduction by measure of GFP in control experiments, was added to each well. Plates were spun at 900g for 90 min at room temperature, then cultured at 37 °C with 5% CO₂-in-air for 4.5 h. At this time, cells were collected, washed extensively with FCS-containing medium and resuspended in the appropriate volume of Hanks Balanced Salt Solution, and between $3\text{--}10 \times 10^3$ progenitors (with an average of 5% transduction) were injected into the indicated recipient mice.

Cell preparation. Organs were collected from mice after they had been killed or by partial splenectomy (see below). To prepare single-cell suspensions, bone marrow cells were extracted as described. Spleen samples were chopped and digested with collagenase type IV (1 mg ml^{–1}) and DNase (20 μ g ml^{–1}) for 20 min at 21 °C, followed by 5 mM EDTA for a further 5 min. Both spleen and bone marrow single-cell suspensions were treated with 0.164 M NH₄Cl for 1 min for red blood cell lysis, and washed. To isolate multiple cell types from spleen and bone marrow samples, a series of magnetic-bead enrichments were performed. Spleen cells were first simultaneously labelled with anti-CD11c beads (Miltenyi) and anti-CD11b biotin for 30 min, washed, and CD11c⁺ dendritic cells (bound) and the remaining non-dendritic-cell fractions (flow) were obtained by magnetic-bead separation. The non-dendritic-cell fraction was then labelled with anti-biotin beads for 10 min by the addition of SA-PerCpCy5.5 after the first 5 min, and subsequently CD11b⁺ myeloid fractions (bound), and non-dendritic-cell and non-myeloid fractions (flow, containing lymphoid cells), were obtained by magnetic-bead separation. For bone marrow samples, the dendritic cell staining and enrichment steps were excluded.

Partial splenectomy and intrasplenic injections. In some experiments, we carried out partial splenectomies or intrasplenic injections that were similar to methods described previously, but with some modifications²⁹. In brief, mice were given 100 μ g kg^{–1} buprenorphine 30 min before surgery, and anaesthetized with isoflurane. Subsequently, the spleen was exposed by incision of the left flank skin and peritoneum and, for partial splenectomy, approximately one-third of the spleen was removed, and the wound sealed with histoacryl (Braun). For intrasplenic injections, 25 μ L of cell suspension was injected just under the capsule of the spleen. In both cases, the remaining spleen was placed back in the peritoneal cavity and peritoneum and skin were sutured separately. Mice were placed in recovery, in a warm environment, and given hydrogel and an additional 100 μ g kg^{–1} Temgesic 6 h later. Within 24 h, wounds healed and the behaviour of the mice returned to normal.

Flow cytometry. Dendritic-cell fractions were stained with antibodies against CD45.2, CD11c, CD8 α , SIRP α and B220, and CD8⁺ dendritic cells, CD11b⁺ dendritic cells and plasmacytoid dendritic cells were obtained by cell sorting. Myeloid-cell fractions that were stained with CD11b were stained further with antibodies against Gr1 and F4/80. The dendritic-cell- and myeloid-cell-depleted fractions were stained with CD19 and B220 antibodies. Cell types were sorted on a FACSAria as CD11c^{hi}B220[–]SIRP α ⁺CD8 α [–] (CD11b⁺ dendritic cells), CD11c⁺B220[–]SIRP α [–]CD8 α ⁺ (CD8⁺ dendritic cells), CD11c⁺B220⁺SIRP α ^{lo} (plasmacytoid dendritic cells), CD11b⁺Gr1^{hi}SSC^{int} (neutrophils), CD11b⁺Gr1^{int}SSC^{lo}F4/80^{lo} (monocytes), and CD19⁺B220⁺ (B cells) (Supplementary Fig. 2). Cells were sorted on the total population and not specifically for GFP-expressing cells to provide ‘filler’ cells for better pellet formation. For sibling analyses and kinetic analyses at day 10, total lineages were sorted rather than each subtype. In this case, cells were stained with

CD11c, CD11b and CD19. CD19[–]CD11c^{hi} dendritic cells, CD19[–]CD11c^{neg/lo}CD11b⁺ myeloid cells and CD19⁺ B cells were sorted as separate populations.

PCR and deep sequencing. Up to 5×10^5 sorted cells from each sample were transferred to different wells of a 96-well PCR plate, with an adjacent well kept empty for a later step. Cells were centrifuged to form pellets, the supernatant removed carefully, and pellets resuspended in 40 μ L PCR lysis buffer (Viagen) containing 200 μ g ml^{–1} proteinase K. Plates were covered with a rubber mat and lysed in a thermocycler at 55 °C for 1 h, and then at 85 °C for 30 min to inactivate proteinase K. Samples were only stored at –20 °C after this step. For the first round of PCR, 160 μ L of PCR reagents containing TopLib 5'-TGCTGCCGTCAACTAGAACAA-3' and BotLib primers 5'-GATCTCGAATCAGGCGCTTA-3' were added to the 40- μ L pellet of all samples. After mixing, 100 μ L was transferred to the adjacent empty well before PCR, to provide the technical replicates to assess barcode detection reliability (Supplementary Figs 1c and 3a)⁹. Plates were sealed and placed in a thermocycler at 94 °C for 5 min, then cycled 30 times at 57.2 °C for 15 s, at 72 °C for 15 s and at 94 °C for 15 s, and then at 72 °C for 10 min. The presence of a 150-bp product was checked for every sample using 2% agarose gel electrophoresis.

In the second round of PCR, a different index primer was used for every sample and technical replicate. To do this, a library of 384 different 82-bp index primers containing unique 8-bp indexes that differed by at least 2 bases was designed, such that homopolymers of more than 2 bp, hairpins and complementary regions with the rest of the primer sequence were absent (sequences available on request). Index primers were ordered desalted, as high-performance liquid chromatography (HPLC) purification resulted in low-level cross-contamination (approximately 0.1%) that could be picked up readily by deep sequencing. Index primers contained a sequence for (in order) the P7 annealing region for the Illumina flow cell, the Illumina sequencing primer, a unique 8-bp index (from 384 such indices) and a 16-bp annealing region to the first round product. A master mix of PCR reagents (24 μ L), which included a common reverse primer that included a P5 annealing region for the Illumina flow cell followed by an annealing region for the first round PCR product 5'-CAAGCAGAAGACGGCATACGAGATTGCTGCGCTCAACTAGAACAA-3', was aliquoted into each well of a 96-well PCR plate. Subsequently, 5 μ L of each index primer, and 1 μ L of the first round PCR product were added to each well. PCR amplification was carried out as described above, and the presence of the expected 224-bp product was checked for each sample by gel electrophoresis.

After a successful second round of PCR, 5 μ L of each of up to 384 samples containing different index primers were pooled, run on an E-Gel (Invitrogen) to isolate the 224-bp product, and used for cluster generation and sequencing on a HiSeq2000 or GAIIx (Illumina). A single-run 50-bp sequencing run was sufficient to read through the index, common annealing region, and the first 15 bp of the barcode required for data analysis.

Bioinformatics. Sequence data were filtered using a customized script in R that searched for perfect matches with one of the index primers, with the common 16-bp annealing region, and with one of the barcodes present in our reference library (Supplementary Methods). Matching against the reference library was important to distinguish true barcodes from sequence errors introduced during amplification and sequencing. In general, approximately 90% of raw sequencing reads passed this filtering step. The average number of filtered reads for all barcodes per sample was approximately 10^5 sequences. As explained in Supplementary Methods, samples subsequently underwent further filtering to ensure that they were representative, and were then normalized for each cell type. Data were tabulated into a matrix containing the fraction of reads for each barcode (that is, progenitor) versus index primer (that is, sample identity) as its elements. Visualization and analysis for clustering and PCA were carried out with MultiExperiment Viewer software³⁰.

LMPP classification and output. To classify individual progenitors for their propensity to generate certain lineages, and given that progenitors were heterogeneous in their clonal output to cell types, we applied an additional normalization on a per progenitor basis; that is, per row. If the proportional output of an LMPP was greater than 2.5% for at least one cell type of a given lineage, the progenitor was considered to have an output towards that lineage. In this way, progenitors were classified into having mono-, bi- or tri-outcome for the B-cell, dendritic-cell and myeloid-cell lineages; that is, we allowed for 7 classes (B, DC, M, B/DC, B/M, M/DC, DC/B/M). **Randomization.** For the sibling analyses we carried out a randomization of barcode identity by keeping the barcode identity in one mouse fixed and randomizing the barcode identity in the other 1,000 times. We then repeated this with the barcode identity fixed in the other mouse. After this, the values were pooled to assess the effect of chance on observing a conserved lineage bias.

28. Gonçalves, M. A. *et al.* Genetic complementation of human muscle cells via directed stem cell fusion. *Mol. Ther.* **16**, 741–748 (2008).
29. Reeves, J. P., Reeves, P. A. & Chin, L. T. Survival surgery: removal of the spleen or thymus. *Curr. Protoc. Immunol.* **2**, 1.10.1–1.10.11 (2001).
30. Saeed, A. I. *et al.* TM4 microarray software suite. *Methods Enzymol.* **411**, 134–193 (2006).

Manipulation of small Rho GTPases is a pathogen-induced process detected by NOD1

A. Marijke Keestra¹, Maria G. Winter¹, Josef J. Auburger¹, Simon P. Fräßle¹, Mariana N. Xavier¹, Sebastian E. Winter¹, Anita Kim¹, Victor Poon¹, Mariëtta M. Ravesloot¹, Julian F. T. Waldenmaier¹, Renée M. Tsois¹, Richard A. Eigenheer² & Andreas J. Bäuml¹

Our innate immune system distinguishes microbes from self by detecting conserved pathogen-associated molecular patterns¹. However, these are produced by all microbes, regardless of their pathogenic potential. To distinguish virulent microbes from those with lower disease-causing potential the innate immune system detects conserved pathogen-induced processes², such as the presence of microbial products in the host cytosol, by mechanisms that are not fully resolved. Here we show that NOD1 senses cytosolic microbial products by monitoring the activation state of small Rho GTPases. Activation of RAC1 and CDC42 by bacterial delivery or ectopic expression of SopE, a virulence factor of the enteric pathogen *Salmonella*, triggered the NOD1 signalling pathway, with consequent RIP2 (also known as RIPK2)-mediated induction of NF- κ B-dependent inflammatory responses. Similarly, activation of the NOD1 signalling pathway by peptidoglycan required RAC1 activity. Furthermore, constitutively active forms of RAC1, CDC42 and RHOA activated the NOD1 signalling pathway. Our data identify the activation of small Rho GTPases as a pathogen-induced process sensed through the NOD1 signalling pathway.

One process that marks pathogens for recognition by the host is the delivery of microbial molecules into the host cell cytosol² (Supplementary Fig. 1). For example, the enteric pathogen *Salmonella enterica* serovar Typhimurium (hereafter termed *S. typhimurium*) uses a type III secretion system (T3SS-1) encoded by *Salmonella* pathogenicity island (SPI) 1 to deliver proteins, termed effectors, into epithelial cells^{3,4}. *S. typhimurium* elicits proinflammatory responses by translocating four T3SS-1 effector proteins, termed SipA, SopE, SopB and SopE2 (refs 5–8). To investigate the mechanism of T3SS-1-dependent NF- κ B activation reported previously^{7–9}, we used human cells transfected with a NF- κ B luciferase reporter construct. A *S. typhimurium* mutant lacking the proinflammatory effector proteins (*sipA sopE sopB sopE2* mutant) was deficient for NF- κ B activation in epithelial HeLa cells ($P < 0.05$; Supplementary Fig. 2a). Inactivation of proinflammatory effector genes resulted in a partial inhibition of NF- κ B activation in HEK293 cells, presumably because endogenous TLR5 expression renders HEK293 cells responsive to stimulation with flagellin (Supplementary Fig. 2b)^{10–12}. Compared to a mutant lacking all four proinflammatory effector proteins, NF- κ B activation was significantly enhanced after infecting host cells with a mutant expressing SopE (*sipA sopB sopE2* mutant) ($P < 0.05$; Supplementary Fig. 2a, b). Cytosolic localization of SopE in the absence of the other bacterial proteins is sufficient for inducing NF- κ B activation⁵, as illustrated by ectopic expression of a green fluorescent protein (GFP)–SopE fusion protein in HEK293 cells (Supplementary Fig. 2c).

In the host cell cytosol, SopE activates RAC1 and CDC42 by serving as a nucleotide-exchange factor that facilitates the transition of the small Rho GTPases from an inactive GDP-bound state to an active GTP-bound state⁵. Transfection of HEK293 cells with plasmids encoding dominant-negative forms of CDC42 (CDC42(DN)) or RAC1 (RAC1(DN))¹³ inhibited NF- κ B activation elicited by ectopic

expression of SopE ($P < 0.05$; Supplementary Fig. 2d, e), which was consistent with previous reports⁵. By contrast, no significant inhibition was observed when HEK293 cells were transfected with a plasmid encoding a dominant-negative form of the small Rho GTPase RHOA (RHOA(DN)). A G168A amino acid substitution in SopE (SopE (G168A)) is known to abrogate its nucleotide-exchange factor activity for RAC1 and CDC42 (ref. 14). Transfection of HEK293 cells with a plasmid encoding GFP–SopE(G168A) no longer resulted in NF- κ B activation (Supplementary Fig. 2f). Collectively, these observations raised the possibility that activation of RAC1 and CDC42 might be a pathogen-induced process detected by host cells as a pattern of pathogenesis¹⁵. However, the identity of the host's pattern-recognition receptor (PRR) that activates NF- κ B when it detects this pattern of pathogenesis remained elusive.

The induction of T3SS-1-dependent inflammatory responses during *S. typhimurium* colitis is largely MYD88-independent¹⁶. One MYD88-independent mechanism by which *S. typhimurium* induces NF- κ B activation is through the induction of NOD1 and NOD2, two cytosolic PRRs that sense cell wall fragments^{12,17,18}. NOD1 and NOD2 interact with the protein kinase RIP2 to mediate NF- κ B activation^{19,20}. Remarkably, treatment of HEK293 cells with SB203580, a pyridinyl imidazole inhibitor of RIP2 activity (Supplementary Fig. 2g)²¹, significantly reduced NF- κ B activation elicited by bacterial delivery of SopE (that is, infection with a *sipA sopB sopE2* mutant) ($P < 0.05$; Fig. 1a). In this model, no SopE-independent NF- κ B activation was observed when cells were infected with non-flagellated *S. typhimurium* strains (Supplementary Fig. 2h), suggesting that flagella were responsible for the background levels of NF- κ B activation elicited by the mutant lacking proinflammatory effector proteins (*sipA sopE sopB sopE2* mutant) (Fig. 1a).

We next investigated whether NOD1 and/or NOD2 were required for SopE-dependent NF- κ B activation by transfecting HEK293 cells with plasmids encoding dominant-negative forms of NOD1 (NOD1(DN)), NOD2 (NOD2(DN)) or RIP2 (RIP2(DN))¹². In control experiments, expression of NOD1(DN) or RIP2(DN) inhibited NF- κ B activation elicited by stimulation with the NOD1 ligand C12-iE-DAP (Supplementary Fig. 3a), whereas expression of NOD2(DN) or RIP2(DN) inhibited NF- κ B activation elicited by stimulation with the NOD2 ligand muramyl dipeptide (MDP) (Supplementary Fig. 3b). Transfection of HEK293 cells with NOD1(DN) or RIP2(DN) inhibited NF- κ B activation elicited by ectopic expression of GFP–SopE ($P < 0.05$; Fig. 1b and Supplementary Fig. 3c). By contrast, transfection with NOD2(DN) did not reduce GFP–SopE-induced NF- κ B activation.

To investigate further whether NOD1 is required for SopE-dependent NF- κ B activation, we silenced *NOD1* expression in HEK293 cells with short interfering RNA (siRNA). In a control experiment, *NOD1*-specific siRNA inhibited NF- κ B activation elicited with C12-iE-DAP, but did not affect responses elicited by MDP or flagellin (Supplementary Fig. 3d). Notably, *NOD1*-specific siRNA significantly ($P < 0.05$) reduced NF- κ B activation elicited through bacterial delivery of SopE

¹Department of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, One Shields Avenue, Davis, California 95616, USA. ²Proteomics Core Facility, UC Davis Genome Center, University of California at Davis, Davis, California 95616, USA.

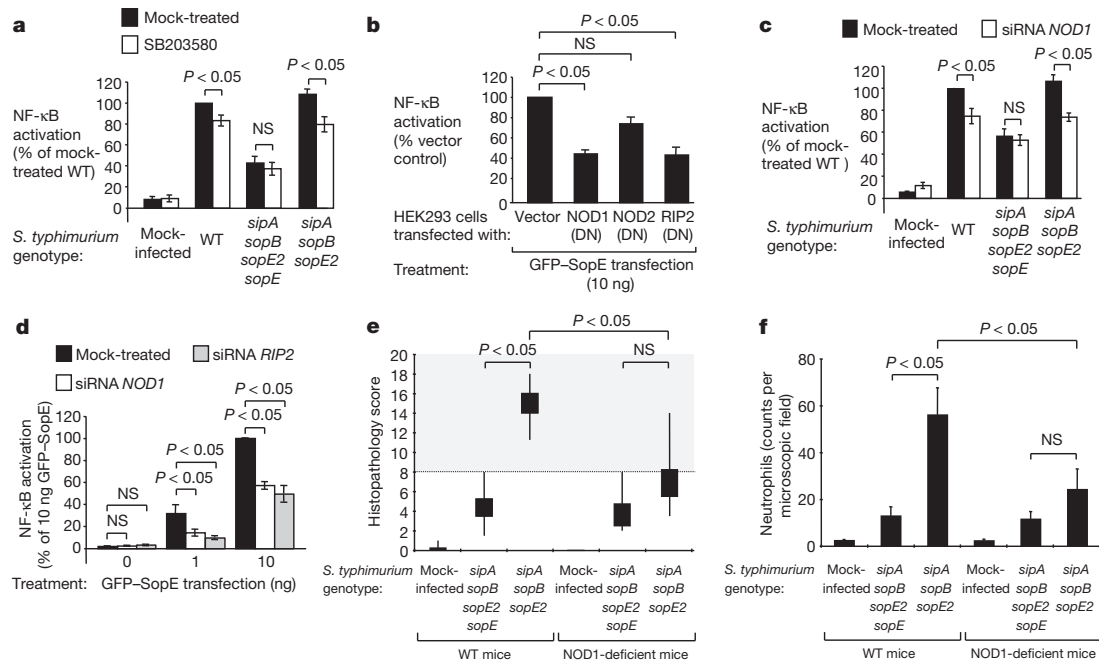


Figure 1 | SopE-induced NF-κB activation requires RIP2 and NOD1.

a–d, HEK293 cells transfected with a NF-κB-luciferase reporter were treated as indicated. In **a** and **c** cells were infected with *S. typhimurium* wild type (WT) or mutants. Mean luciferase activity \pm s.e.m., $n = 3$. **e**, **f**, Streptomycin-pretreated NOD1-deficient mice and wild-type littermates (for n see Supplementary Fig.

5b) were infected and the caecum collected after 24 h. Average histopathology is shown in panel **e**. Whisker plots represent the second and third quartiles (boxes) and the first and fourth quartiles (lines). **f** shows average number of neutrophils per microscopic field \pm s.e.m. NS, not significant.

by flagellated (Fig. 1c) or non-flagellated (Supplementary Fig. 3e) *S. typhimurium* strains. To investigate further whether NOD1 and RIP2 were required for SopE-dependent NF-κB activation, we silenced *NOD1* and *RIP2* expression in cells ectopically expressing SopE. In a control experiment, *RIP2*-specific siRNA inhibited NF-κB activation elicited by C12-iE-DAP and MDP, but not by flagellin (Supplementary Fig. 3d). Remarkably, *NOD1*-specific siRNA and *RIP2*-specific siRNA inhibited NF-κB activation elicited by ectopic expression of GFP-SopE ($P < 0.05$; Fig. 1d). Inhibition of NF-κB activation by NOD1(DN), *RIP2*(DN), *NOD1*-specific siRNA or *RIP2*-specific siRNA could be overcome when large quantities (100 ng) of the plasmid encoding GFP-SopE were used for transfection (Supplementary Fig. 4a, b), which resulted in cells lifting off from wells (Supplementary Fig. 4c). These effects might explain why a contribution of *RIP2* to SopE-induced host cell responses was missed in a previous study⁷. Collectively, our data indicate that SopE-mediated NF-κB activation requires both NOD1 and *RIP2* activity.

To test the biological relevance of our observations, we determined the role of NOD1 in SopE-mediated inflammation *in vivo* using the mouse colitis model of *S. typhimurium* infection²². To restrict our analysis to SopE-mediated mechanisms of inflammation, we compared a *S. typhimurium* strain producing only SopE (*sipA sopB sopE2* mutant) with an isogenic SopE-deficient mutant (*sipA sopE sopB sopE2* mutant). Remarkably, the SopE-proficient *S. typhimurium* strain triggered caecal inflammation in wild-type littermates, but not in NOD1-deficient mice (Fig. 1e, f and Supplementary Fig. 5a). By contrast, a SopE-deficient mutant did not elicit marked caecal inflammation, although the strain was recovered from the caecum in numbers similar to those of the SopE-proficient strain (Supplementary Fig. 5b). Thus, a SopE-proficient *S. typhimurium* strain required NOD1 for eliciting intestinal inflammation *in vivo*.

Because SopE-mediated NF-κB activation was both RAC1-dependent (Supplementary Fig. 2d) and NOD1-dependent (Fig. 1), we investigated whether NOD1 senses the activation state of RAC1 (Fig. 2). HEK293 cells were transfected with a plasmid encoding a RAC1 derivative carrying an amino acid substitution (Q61L) that abrogates its

GTPase activity, resulting in a constitutively active, GTP-bound form of RAC1 (RAC1(CA))¹³. Expression of RAC1(CA) induced NF-κB activation, which could be blunted by expressing NOD1(DN) or *RIP2*(DN) (Fig. 2a and Supplementary Fig. 6a). Similarly, introducing siRNA specific for *NOD1* or *RIP2* blunted NF-κB activation induced by RAC1(CA) (Fig. 2b). Notably, constitutively active forms of CDC24 (CDC42(CA)) and RHOA (RHOA(CA)) also induced NF-κB activation in a NOD1- and *RIP2*-dependent manner (Fig. 2c, d and Supplementary Fig. 6a), suggesting that the NOD1 signalling pathway senses the activation of several small Rho GTPases. An intact membrane anchor of active RAC1 was important for NF-κB activation, because a GTP-bound form of RAC1 lacking its prenyl group (RAC1((CAp))²³ did not activate NF-κB, and a dominant-negative form of RAC1 lacking its prenyl-group (RAC1(DNp)) was no longer capable of inhibiting SopE-induced NF-κB activation (Fig. 2e and Supplementary Fig. 6b). Together, these data suggested that NOD1 detects bacterial-induced activation of small Rho GTPases in host cells.

Because NOD1 is known to sense the presence of Gram-negative peptidoglycan in the cytosol²⁰, we wanted to investigate whether detecting this pattern of pathogenesis required small Rho GTPases. When we transfected HEK293 cells with increasing amounts of plasmid encoding GFP-SopE and stimulated them with different concentrations of C12-iE-DAP, the effects on NF-κB activation were additive, not synergistic (Fig. 2f). Similarly, RAC1(CA) and C12-iE-DAP did not synergize in inducing NF-κB activation (Supplementary Fig. 7). Remarkably, NF-κB activation elicited by treatment of cells with C12-iE-DAP was blunted when cells were transfected with RAC1(DN), but not with CDC42(DN) (Fig. 2g). By contrast, NF-κB activation elicited by treatment of cells with flagellin was not affected by transfection with RAC1(DN) or CDC42(DN). In summary, our data indicate that NOD1 detects peptidoglycan as a pattern of pathogenesis by sensing the activation state of RAC1.

To investigate a possible interaction of NOD1 with SopE and small Rho GTPases we performed confocal microscopy with HEK293 cells ectopically expressing haemagglutinin-tagged SopE (SopE-HA) (Fig. 3). Expression of SopE-HA resulted in the formation of membrane ruffles

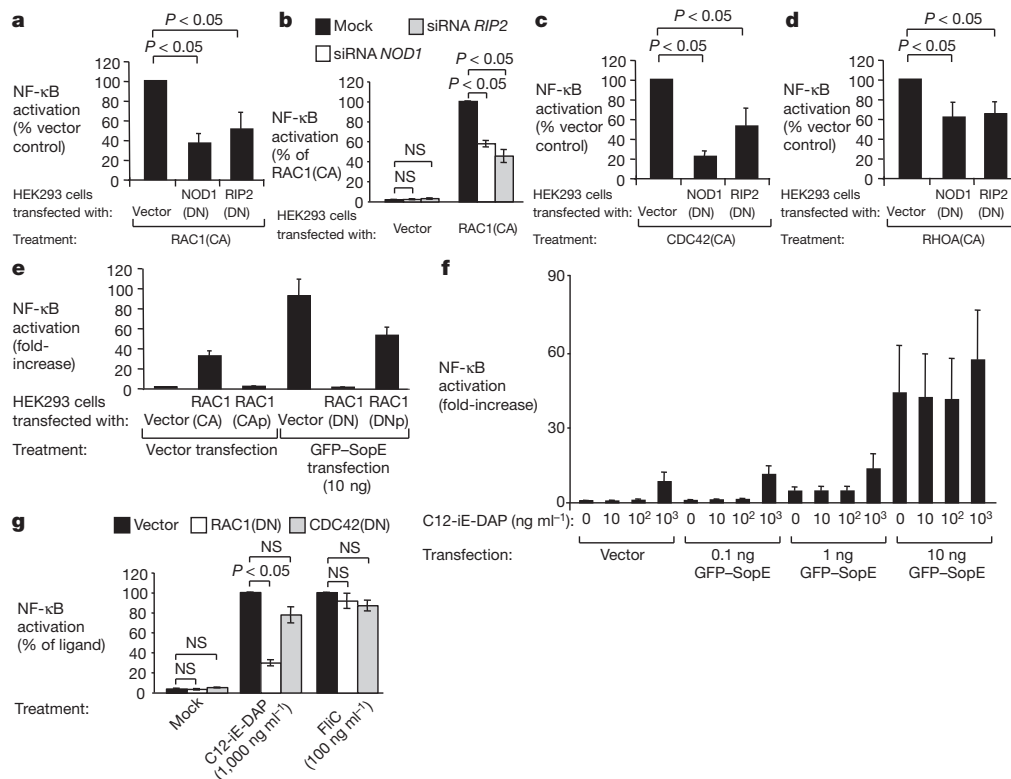


Figure 2 | NOD1 senses the activation state of small Rho GTPases. HEK293 cells were transfected as indicated, together with a NF-κB luciferase reporter. **a–e**, Luciferase activity was measured 48 h after transfection. Data are presented as mean \pm s.e.m. from at least three independent experiments. Brackets

indicate the significance of differences. **f, g**, Cells were stimulated with C12-iE-DAP or flagellin (FliC) as indicated and luciferase activity was measured 5 h thereafter.

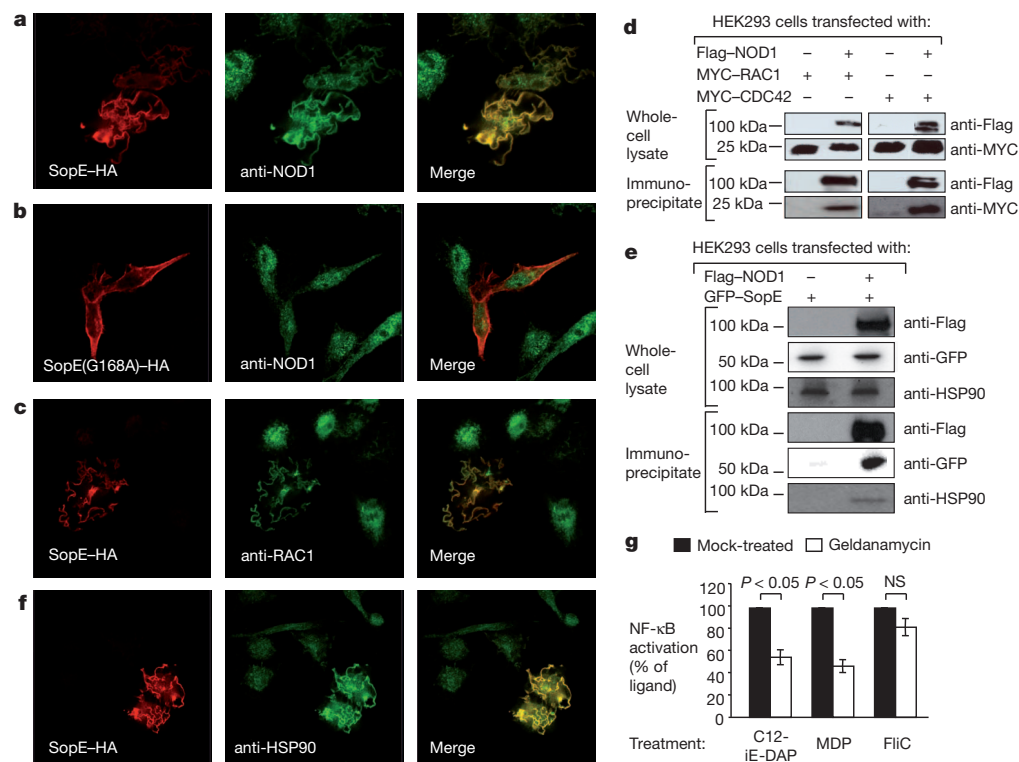


Figure 3 | SopE forms a multiprotein complex with RAC1, CDC42, NOD1 and HSP90. **a–c, f**, Ectopically expressed SopE-HA (**a, c, f**), SopE(G168A)-HA (**b**) and endogenous NOD1 (**a, b**), RAC1 (**c**) and HSP90 (**f**) were detected by confocal microscopy. **d, e**, Whole-cell lysates and anti-Flag immunoprecipitates were separated by SDS-polyacrylamide gel

electrophoresis and subjected to immunoblotting. **g**, HEK293 cells transfected with a NF-κB luciferase reporter were pretreated with the HSP90 inhibitor geldanamycin (100 nm). Data are presented as mean \pm s.e.m. of at least three independent experiments. Brackets indicate the significance of differences.

and the recruitment of endogenous NOD1, which was detected with an anti-NOD1 antibody (Fig. 3a). By contrast, a SopE derivative carrying an amino acid substitution that prevents activation of RAC1 and CDC42 (SopE(G168A)-HA) did not induce the formation of membrane ruffles and no co-localization with endogenous NOD1 was observed (Fig. 3b). Endogenous RAC1, which was detected with an anti-RAC1 antibody, also co-localized with SopE-HA in membrane ruffles (Fig. 3c). When HEK293 cells expressing Flag-tagged NOD1 (Flag-NOD1) were co-transfected with plasmids encoding either MYC-tagged RAC1 (MYC-RAC1) or MYC-CDC42, a monoclonal anti-Flag antibody co-immunoprecipitated MYC-RAC1 and MYC-CDC42 (Fig. 3d), suggesting that NOD1 is present in a protein complex containing these small Rho GTPases. Further analysis of immunoprecipitates by mass spectrometry identified HSP90 (also known as HSP90AB1) as a possible component of NOD1-containing multiprotein complexes (Supplementary Table 1). To confirm this interaction, cells expressing Flag-NOD1 and GFP-SopE were immunoprecipitated with a monoclonal anti-Flag antibody. GFP-SopE and endogenous cellular HSP90 were detected in the immunoprecipitate (Fig. 3e). Furthermore, endogenous HSP90 co-localized with SopE-HA in membrane ruffles (Fig. 3f). To investigate whether the presence of HSP90 is required for signalling, we abolished HSP90 activity in HEK293 cells by treatment with the specific HSP90 inhibitor geldanamycin. Geldanamycin treatment inhibited NF- κ B activation elicited by stimulation with C12-iE-DAP and MDP, but did not significantly reduce NF- κ B activation induced by flagellin (Fig. 3g). Furthermore, treatment of HEK293 cells with geldanamycin significantly reduced NF- κ B activation elicited by bacterial delivery of SopE (that is, infection with a *sipA sopB sopE2* mutant) ($P < 0.05$; Supplementary Fig. 8).

We conclude that a multiprotein complex composed of small Rho GTPases, HSP90 and NOD1 responds to pathogens gaining cytosolic access by activating NF- κ B (Supplementary Fig. 1). Escape of pathogens (for example, *Shigella flexneri*) from the phagosome introduces peptidoglycan fragments into the cytosol²⁰ and our data indicate that a complex composed of active RAC1, HSP90 and NOD1 detects this pattern of pathogenesis. In addition, pathogens commonly manipulate the actin cytoskeleton of host cells by injecting proteins into the cytosol that alter the activity of small Rho GTPases^{15,24}. Here it is shown that activation of small Rho GTPases by a bacterial virulence factor is a second pattern of pathogenesis detected through the NOD1 signalling pathway in mammalian cells. Alternatively, manipulation of small Rho GTPases by effector proteins could be viewed as a mechanism through which pathogens manipulate host responses to gain a growth advantage²⁵. Effector-triggered immune responses of invertebrates might represent an interesting parallel, because activation of the small Rho GTPase RAC2 in *Drosophila* by a bacterial toxin leads to activation of the innate immune adaptor IMD, which shares homology to the mammalian RIP1 (RIPK1) protein²⁶. RIP1 and RIP2 are important activators of NF- κ B in response to cellular stress in mammals. RIP2 binds the caspase-recruitment domains of NOD1 and NOD2 to initiate NF- κ B activation²⁷. Our data indicate that NOD1 interacts with proteins in addition to RIP2, because this PRR was present in a multiprotein complex that contained a bacterial effector (SopE), small Rho GTPases and HSP90. The term 'nodosome' has been proposed for mammalian NOD1- and/or NOD2-containing protein complexes functioning in NF- κ B activation²⁸. RHOA has previously been implicated in NOD1-dependent NF- κ B activation induced by the *S. flexneri* effector protein OspB; however, this small Rho GTPase was proposed to act downstream of NOD1 (ref. 29). The finding that NOD1 senses the activation state of small Rho GTPases acting upstream in the signalling cascade (Supplementary Fig. 1) considerably changes our understanding of the mechanisms that trigger this signalling pathway.

METHODS SUMMARY

Transfections. Transfections were performed using FuGENE HD (Roche) according to the manufacturer's instructions. Forty-eight hours after transfection,

cells were lysed either without any treatment, or infected with the indicated bacterial strains or the NOD1, NOD2 and TLR5 ligands with or without pretreatment (30 min) of the cells with the inhibitors (10 μ M SB203580 and 100 nM geldanamycin; InvivoGen). The cells were infected with the indicated *S. typhimurium* strains (10^6 colony-forming units (c.f.u.) ml⁻¹) for 1 h, after which the cells were washed with Dulbecco's phosphate-buffered saline and incubated at 37 °C for an additional 4 h in the presence of Dulbecco's modified Eagle's medium (DMEM) containing 10% FCS. After 5 h of treatment the cells were lysed and analysed for β -galactosidase activity and luciferase activity¹².

Animal experiments. All mouse experiments were approved by the Institutional Animal Care and Use Committees at the University of California, Davis. *Nod1*^{+/-} *Nod2*^{+/-} (wild-type littermates, $n = 8$) and *Nod1*^{-/-} *Nod2*^{+/-} (NOD1-deficient, $n = 8$) streptomycin-pretreated mice were orally inoculated with 0.1 ml sterile Luria-Bertani (LB) broth or 1×10^9 c.f.u. (in 0.1 ml LB broth) of a *S. typhimurium sipA sopE sopB sopE2* or a *sipA sopB sopE2* mutant. Caecal contents and tissues were collected 24 h after infection for further analysis.

Fluorescence microscopy. HeLa cells transiently transfected with expression plasmids were fixed with 3% paraformaldehyde as described previously¹². Cells were incubated with mouse anti-HA (Covance) and rabbit anti-NOD1, rabbit anti-HSP90 (Cell Signaling) or rabbit anti-RAC1 (Fisher). Cells were incubated with Alexa fluor 647 goat anti-mouse IgG and Alexa fluor 488 goat anti-rabbit IgG (Invitrogen). Samples were embedded in FluorSave (Calbiochem) and analysed with a LSM700 Confocal Microscope.

Statistical analysis. Statistical differences were calculated using a paired Student's *t*-test. To determine the statistically significant differences in the total histopathology scores, a Mann-Whitney *U*-test was used. A two-tailed $P < 0.05$ was considered to be significant.

Full Methods and any associated references are available in the online version of the paper.

Received 3 February 2012; accepted 19 February 2013.

Published online 31 March 2013.

- Janeway, C. A. Jr Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.* **54**, 1–13 (1989).
- Vance, R. E., Isberg, R. R. & Portnoy, D. A. Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system. *Cell Host Microbe* **6**, 10–21 (2009).
- Fu, Y. & Galan, J. E. The *Salmonella typhimurium* tyrosine phosphatase SptP is translocated into host cells and disrupts the actin cytoskeleton. *Mol. Microbiol.* **27**, 359–368 (1998).
- Zhou, D., Mooseker, M. S. & Galan, J. E. An invasion-associated *Salmonella* protein modulates the actin-bundling activity of platin. *Proc. Natl Acad. Sci. USA* **96**, 10176–10181 (1999).
- Hardt, W. D., Chen, L. M., Schuebel, K. E., Bustelo, X. R. & Galan, J. E. *S. typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* **93**, 815–826 (1998).
- Hapfelmeier, S. *et al.* Role of the *Salmonella* pathogenicity island 1 effector proteins SipA, SopB, SopE, and SopE2 in *Salmonella enterica* subspecies 1 serovar Typhimurium colitis in streptomycin-pretreated mice. *Infect. Immun.* **72**, 795–809 (2004).
- Bruno, V. M. *et al.* *Salmonella* Typhimurium type III secretion effectors stimulate innate immune responses in cultured epithelial cells. *PLoS Pathog.* **5**, e1000538 (2009).
- Figueiredo, J. F. *et al.* *Salmonella enterica* Typhimurium SipA induces CXCL1 chemokine expression through p38MAPK and JUN pathways. *Microbes Infect.* **11**, 302–310 (2009).
- Hobbie, S., Chen, L. M., Davis, R. J. & Galan, J. E. Involvement of mitogen-activated protein kinase pathways in the nuclear responses and cytokine production induced by *Salmonella typhimurium* in cultured intestinal epithelial cells. *J. Immunol.* **159**, 5550–5559 (1997).
- Smith, M. F. Jr *et al.* Toll-like receptor (TLR) 2 and TLR5, but not TLR4, are required for *Helicobacter pylori*-induced NF- κ B activation and chemokine expression by epithelial cells. *J. Biol. Chem.* **278**, 32552–32560 (2003).
- Lu, W. *et al.* Cutting edge: enhanced pulmonary clearance of *Pseudomonas aeruginosa* by Muc1 knockout mice. *J. Immunol.* **176**, 3890–3894 (2006).
- Kestra, A. M. *et al.* A *Salmonella* virulence factor activates the NOD1/NOD2 signaling pathway. *mBio* **2**, e00266–00211 (2011).
- Coso, O. A. *et al.* The small GTP-binding proteins Rac1 and Cdc42 regulate the activity of the JNK/SAPK signaling pathway. *Cell* **81**, 1137–1146 (1995).
- Schlumberger, M. C. *et al.* Amino acids of the bacterial toxin SopE involved in G nucleotide exchange on Cdc42. *J. Biol. Chem.* **278**, 27149–27159 (2003).
- Müller, A. J., Hoffmann, C. & Hardt, W. D. Caspase-1 activation via Rho GTPases: a common theme in mucosal infections? *PLoS Pathog.* **6**, e1000795 (2010).
- Hapfelmeier, S. *et al.* The *Salmonella* pathogenicity island (SPI)-2 and SPI-1 type III secretion systems allow *Salmonella* serovar typhimurium to trigger colitis via MyD88-dependent and MyD88-independent mechanisms. *J. Immunol.* **174**, 1675–1685 (2005).
- Le Bourhis, L. *et al.* Role of Nod1 in mucosal dendritic cells during *Salmonella* pathogenicity island 1-independent *Salmonella enterica* serovar Typhimurium infection. *Infect. Immun.* **77**, 4480–4486 (2009).

18. Geddes, K. *et al.* Nod1 and Nod2 regulation of inflammation in the *Salmonella colitis* model. *Infect. Immun.* **78**, 5107–5115 (2010).
19. Inohara, N. *et al.* An induced proximity model for NF- κ B activation in the Nod1/RICK and RIP signaling pathways. *J. Biol. Chem.* **275**, 27823–27831 (2000).
20. Girardin, S. E. *et al.* CARD4/Nod1 mediates NF- κ B and JNK activation by invasive *Shigella flexneri*. *EMBO Rep.* **2**, 736–742 (2001).
21. Argast, G. M., Fausto, N. & Campbell, J. S. Inhibition of RIP2/RICK/CARDIAK activity by pyridinyl imidazole inhibitors of p38 MAPK. *Mol. Cell. Biochem.* **268**, 129–140 (2005).
22. Barthel, M. *et al.* Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar Typhimurium colitis model that allows analysis of both pathogen and host. *Infect. Immun.* **71**, 2839–2858 (2003).
23. Wong, K. W., Mohammadi, S. & Isberg, R. R. The polybasic region of Rac1 modulates bacterial uptake independently of self-association and membrane targeting. *J. Biol. Chem.* **283**, 35954–35965 (2008).
24. Aktories, K., Schmidt, G. & Just, I. Rho GTPases as targets of bacterial protein toxins. *Biol. Chem.* **381**, 421–426 (2000).
25. Lopez, C. A. *et al.* Phage-mediated acquisition of a type III secreted effector protein boosts growth of *Salmonella* by nitrate respiration. *mBio* **3**, e00143–12 (2012).
26. Boyer, L. *et al.* Pathogen-derived effectors trigger protective immunity via activation of the Rac2 enzyme and the IMD or Rip kinase signaling pathway. *Immunity* **35**, 536–549 (2011).
27. Meylan, E. & Tschopp, J. The RIP kinases: crucial integrators of cellular stress. *Trends Biochem. Sci.* **30**, 151–159 (2005).
28. Tattoli, I., Travassos, L. H., Carneiro, L. A., Magalhaes, J. G. & Girardin, S. E. The Nodosome: Nod1 and Nod2 control bacterial infections and inflammation. *Semin. Immunopathol.* **29**, 289–301 (2007).
29. Fukazawa, A. *et al.* GEF-H1 mediated control of NOD1 dependent NF- κ B activation by *Shigella* effectors. *PLoS Pathogens* **4**, e1000228 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank S.-P. Nuccio for providing PCR primers for the construction of the bacterial strains. This work was supported by Public Health Service Grants AI044170 and AI076246. A.M.K. is supported by the American Heart Association Grant 12SDG12220022.

Author Contributions A.M.K. contributed to the experimental design, performed experiments and contributed to Figs 1b–f, 2a–e, g, 3a–c, f, g and Supplementary Figs 1, 2d, 3a, b, d, 4a–c, 7 and 8. M.G.W. performed experiments, constructed bacterial strains and contributed to Figs 1a, e, f, 3d, g and Supplementary Figs 2a, b, e, h, g, e, 4c, 5b and 8. J.J.A. constructed expression plasmids and contributed to Figs 2f and 3c and Supplementary Fig. 7. S.P.F. constructed expression plasmids and contributed to Fig. 2g and Supplementary Figs 2f, 3c and 6a, b. M.N.X. contributed to Fig. 1e, f and Supplementary Fig. 5a. S.E.W. constructed bacterial strains, contributed to Supplementary Fig. 5b and critically read the manuscript. A.K. constructed expression plasmids and contributed to Supplementary Figs 2c and 7. V.P. constructed bacterial strains. M.M.R. contributed to Supplementary Fig. 2d. J.F.T.W. constructed expression plasmids. R.A.E. performed mass spectrometry. A.M.K., A.J.B. and R.M.T. provided financial support for the study and contributed to the experimental design. A.M.K. and A.J.B. were responsible for the overall study design and for writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.J.B. (ajbaumler@ucdavis.edu).

METHODS

Bacterial strains, tissue culture cells and culture conditions. The bacterial strains used in this study are listed in Supplementary Table 2. *E. coli* and *S. typhimurium* strains were routinely cultured aerobically at 37 °C in Luria-Bertani (LB) broth or on LB agar plates supplemented with antibiotics. HeLa57A and HEK293 cell lines have been described previously¹² and were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% FCS at 37 °C in a 5% CO₂ atmosphere.

Animal experiments. All mouse experiments were approved by the Institutional Animal Care and Use Committees at the University of California, Davis. *Nod1*^{+/-} *Nod2*^{+/-} and *Nod1*^{-/-} *Nod2*^{+/-} mice on a C57BL/6 background were bred and housed at the Center for Laboratory Animal Science at the University of California Davis. DNA was isolated from mouse tails with the DNeasy Blood & Tissue kit (Qiagen) and used as a template for genotyping. Primers for genotyping are listed in Supplementary Table 3. Streptomycin (Sigma; 20 mg per mouse)-pretreated mice were orally inoculated with 0.1 ml of sterile LB broth or 1×10^9 colony-forming units (c.f.u.) (in 0.1 ml of LB broth) of an *S. typhimurium* *sipA sopE sopB sopE2* mutant or a *sipA sopB sopE2* mutant. At 24 h after infection, mice were euthanized and tissue samples were collected. Caecal contents were collected in PBS buffer and serial dilutions were plated on LB agar containing the appropriate antibiotics to determine bacterial numbers. Formalin-fixed, hematoxylin and eosin-stained caecal tissue sections were blinded for evaluation by a veterinary pathologist. The following pathological changes were scored as follows: (1) neutrophil infiltration; (2) infiltration by mononuclear cells; (3) submucosal oedema; (4) epithelial damage; and (5) inflammatory exudate. The pathological changes were scored on a scale from 0 to 4 as follows: 0, no changes; 1, detectable; 2, mild; 3, moderate; 4, severe. Neutrophil counts were determined by high-magnification ($\times 400$) microscopy and numbers were averaged from ten microscopic fields for each animal. Images were taken using an Olympus BX41 microscope.

Construction of plasmids. Plasmids and primers used in this study are listed in Supplementary Tables 2 and 3. Standard cloning techniques were performed as described previously³⁰. A DNA fragment upstream of the *sipA*-coding sequence was amplified from *S. typhimurium*, cloned into pCR2.1 using the TOPO-TA cloning kit (Life Technologies) and subcloned into pRDH10 using EcoRI, yielding pSW244. To facilitate transduction of the unmarked $\Delta sipA$ mutation, pSW244 was introduced in this locus in the ZA10 (*S. typhimurium sipA* mutant) chromosome by conjugation with S17-1 λ pir as the donor strain, creating SW974. The *sopE* gene was PCR amplified from M30 and cloned into the mammalian expression vectors pEGFP-C1 and pCMV-HA (BD Biosciences Clontech). The small GTPases RAC1, CDC42 and RHOA were PCR amplified from complementary DNA prepared from HeLa57A cells. SopE(G168A) was PCR amplified from pEGFP-SopE. The dominant-negative forms of RAC1 (RAC1(DN)), CDC42 (CDC42(DN)) and RHOA (RHOA(DN)) are mutated in the GTP-binding site at positions 17, 17 and 19, respectively, from a threonine to an asparagine¹³. The constitutively active forms of RAC1 (RAC1(CA)), CDC42 (CDC42(CA)) and RHOA (RHOA(CA)) are mutated in the GTP-hydrolysis site at positions 61, 61 and 63, respectively, from a glutamine to a leucine¹³. The prenylation mutants of RAC1 are mutated at position 189, where a cysteine is replaced by a serine²³. SopE(G168A) and the dominant-negative and constitutively active mutant forms of RAC1, CDC42 and RHOA were engineered by overlap extension PCR. The two DNA fragments needed for the construction of these mutant forms were amplified by PCR, gel-purified and then fused in one PCR reaction before cloning into pEGFP-C1, pCMV-HA and pCMV-MYC. All constructs were verified by DNA sequencing (SeqWright).

Transfections. Transfections were performed using FuGENE HD (Roche) according to the manufacturer's instructions. Forty-eight hours after transfection, cells were lysed either without any treatment, or infected with the indicated bacterial strains or the NOD1, NOD2 and TLR5 ligands with or without pretreatment (30 min) of the cells with the inhibitors SB203580 (10 μ M) and geldanamycin (100 nM) (InvivoGen). The cells were infected with the indicated *S. typhimurium* strains (10^6 c.f.u. ml⁻¹) for 1 h, after which the cells were washed with DPBS and incubated at 37 °C for an additional 4 h in the presence of DMEM containing 10% FCS. After 5 h of treatment the cells were lysed and analysed for β -galactosidase activity and luciferase activity¹².

Generalized phage transduction. Phage P22 HT int-105 was used for transductions³¹. A lysate of ZA21 was used to separately transduce the *sopB::MudJ* and *sopE2::pSB1039* mutations into SL1344, creating SW798 and SW800, respectively. The $\Delta sipA::pSW244$ mutation from SW974 was separately introduced into SL1344 as well as the $\Delta sopE$ mutant SW976 by transduction, and subsequent sucrose selection³² was performed to create SW1007 and SW1009, respectively. To construct SW973 and SW1008, a P22 phage lysate of SW798 was used to transduce the *sopB::MudJ* mutation into SW1009 and SW1007, respectively. A P22 lysate of SW800 was used to transduce the *sopE2::pSB1039* mutation into SW973 and SW1008, generating SW868 and SW867, respectively.

To generate strains AMK457 and AMK456, the $\Delta fliC::pSPN29$ mutation from SPN305 was separately introduced into SW1008 and SW973 by transduction and subsequent sucrose selection was performed to remove pSPN29. The *fljB::MudCm* mutation from SPN287 was then transduced into AMK457 and AMK456 to create AMK461 and AMK460, respectively. A P22 lysate of SW800 was used to introduce the *sopE2::pSB1039* mutation into the AMK461 and AMK460 chromosome, thus creating AMK465 and AMK464, respectively.

Co-immunoprecipitation and western blotting. Transfected HEK293 cells were lysed and co-immunoprecipitation assays were performed with protein G beads (Invitrogen) coated with anti-Flag M2 antibody (Sigma)¹². Whole-cell lysates and immunoprecipitated proteins were analysed for protein expression by western blot using antibodies raised against rabbit anti-GFP, mouse anti-Flag (Sigma), mouse anti-HA (Covance), rabbit anti-MYC, rabbit anti-HSP90 or anti-tubulin (Cell Signaling Technology)¹².

Fluorescence microscopy. HeLa cells transiently transfected with expression plasmids were fixed with 3% paraformaldehyde as described previously¹². Cells were incubated with mouse anti-HA (Covance) and rabbit anti-NOD1, rabbit anti-HSP90 (Cell Signaling) or rabbit anti-RAC1 (Fisher). Cells were incubated with Alexa fluor 647 goat anti-mouse IgG and Alexa fluor 488 goat anti-rabbit IgG (Invitrogen). Samples were embedded in Fluorsave (Calbiochem) and analysed with a LSM700 Confocal Microscope.

Mass spectrometry of NOD1-Flag-interacting proteins. NOD1-Flag-interacting proteins bound to Protein G beads were digested directly with trypsin (Promega, sequencing grade) overnight in 50 mM ammonium bicarbonate, pH 8. Extracted peptides were analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) on a Thermo-Finnigan LTQ with Michrom Paradigm LC and CTC Pal autosampler. MS and MS/MS spectra were acquired using a top 10 method with an MS survey scan obtained for the *m/z* range 375–1,300. Tandem mass spectra were extracted with Xcalibur version 2.0.7. All MS/MS samples were analysed using X!Tandem (The GPM, thegpm.org; version TORNADO (2010.01.01.4)). X!Tandem was set up to search the uniprottaxon_20100804_YNtL3Y human database (44,512 entries) appended to the cRAP database of common contaminants; samples were also searched with an identical but reversed database to calculate false-discovery rate (FDR). Scaffold (version Scaffold_3_00_08, Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm³³. Protein identifications were accepted if they could be established at greater than 95.0% probability. These parameters yielded a protein FDR of 0.6% and a peptide FDR of 0.1% for identification of interacting proteins.

Statistical analysis. Statistical differences were calculated using a paired Student's *t*-test. To determine the statistically significant differences in the total histopathology scores, a Mann-Whitney *U*-test was used. A two-tailed *P* < 0.05 was considered to be significant.

30. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harb. Lab. Press, 1989).
31. Schmieger, H. Phage P22-mutants with increased or decreased transduction abilities. *Mol. Gen. Genet.* **119**, 75–88 (1972).
32. Lawes, M. & Maloy, S. MudSacl, a transposon with strong selectable and counterselectable markers: use for rapid mapping of chromosomal mutations in *Salmonella typhimurium*. *J. Bacteriol.* **177**, 1383–1387 (1995).
33. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).

Succinate is an inflammatory signal that induces IL-1 β through HIF-1 α

G. M. Tannahill¹, A. M. Curtis¹, J. Adamik², E. M. Palsson-McDermott¹, A. F. McGettrick¹, G. Goel³, C. Frezza^{4,5}, N. J. Bernard¹, B. Kelly¹, N. H. Foley¹, L. Zheng⁴, A. Gardet⁶, Z. Tong⁷, S. S. Jany¹, S. C. Corr¹, M. Haneklaus¹, B. E. Caffrey⁸, K. Pierce⁶, S. Walmsley⁹, F. C. Beasley¹⁰, E. Cummins¹¹, V. Nizet¹⁰, M. Whyte⁹, C. T. Taylor¹¹, H. Lin⁷, S. L. Masters¹², E. Gottlieb⁴, V. P. Kelly¹, C. Clish⁶, P. E. Auron^{2*}, R. J. Xavier^{3,5*} & L. A. J. O'Neill¹

Macrophages activated by the Gram-negative bacterial product lipopolysaccharide switch their core metabolism from oxidative phosphorylation to glycolysis¹. Here we show that inhibition of glycolysis with 2-deoxyglucose suppresses lipopolysaccharide-induced interleukin-1 β but not tumour-necrosis factor- α in mouse macrophages. A comprehensive metabolic map of lipopolysaccharide-activated macrophages shows upregulation of glycolytic and down-regulation of mitochondrial genes, which correlates directly with the expression profiles of altered metabolites. Lipopolysaccharide strongly increases the levels of the tricarboxylic acid cycle intermediate succinate. Glutamine-dependent anaplerosis is the principal source of succinate, although the 'GABA (γ -aminobutyric acid) shunt' pathway also has a role. Lipopolysaccharide-induced succinate stabilizes hypoxia-inducible factor-1 α , an effect that is inhibited by 2-deoxyglucose, with interleukin-1 β as an important target. Lipopolysaccharide also increases succinylation of several proteins. We therefore identify succinate as a metabolite in innate immune signalling, which enhances interleukin-1 β production during inflammation.

Activation of Toll-like receptors, notably Toll-like receptor 4, leads to a switch from oxidative phosphorylation to glycolysis in immune cells^{1,2}, similar to that occurring in tumours. In bone-marrow-derived macrophages (BMDMs), 2-deoxyglucose (2DG) specifically inhibits lipopolysaccharide (LPS)- and *Bordetella pertussis*-induced interleukin-1 β (*Il1b*) transcription, but not tumour-necrosis factor- α (*Tnf*) (Fig. 1a, b) or interleukin-6 (*Il6*) (Supplementary Figs 1 and 2) transcription. 2DG had no effect on invasion and growth of the bacteria in BMDMs (Supplementary Fig. 3). The inhibitory effect of 2DG on LPS-induced IL-1 β was evident *in vivo*. Inhibition of TNF- α *in vivo* was also evident, most probably because of an IL-1 β -dependency on induction of TNF- α (Fig. 1c). There was no effect on the induction of IL-6 *in vivo* (Supplementary Fig. 4).

Supplementary Fig. 5 lists LPS-regulated genes affected by 2DG, including *Il1b*. Several hypoxia-inducible factor-1 α (HIF-1 α) targets were upregulated by LPS and downregulated with 2DG, including ankyrin repeat domain 37, lysyl oxidase and activating transcription factor 3.

LPS-induced HIF-1 α protein but not messenger RNA expression in BMDMs was inhibited by 2DG (Fig. 2a and Supplementary Fig. 6). To examine a direct functional relationship between HIF-1 α and IL-1 β , we found LPS-induced IL-1 β protein expression was markedly increased under hypoxia (Fig. 2b), TNF- α was not affected and, as previously shown³, IL-6 expression was inhibited (Supplementary Fig. 7).

The prolyl hydroxylase (PHD) inhibitor dimethyloxallyl glycine, which stabilizes HIF-1 α protein, also boosted LPS-induced *Il1b* mRNA (Supplementary Fig. 8). Conversely, pre-treating LPS-stimulated BMDMs with a cell-permeable α -ketoglutarate (α -KG) derivative, which increases PHD activity, depleting HIF-1 α , significantly reduced LPS-induced *Il1b* mRNA (Fig. 2c). α -KG inhibited expression of both LPS-induced HIF-1 α and IL-1 β protein in a dose-dependent manner

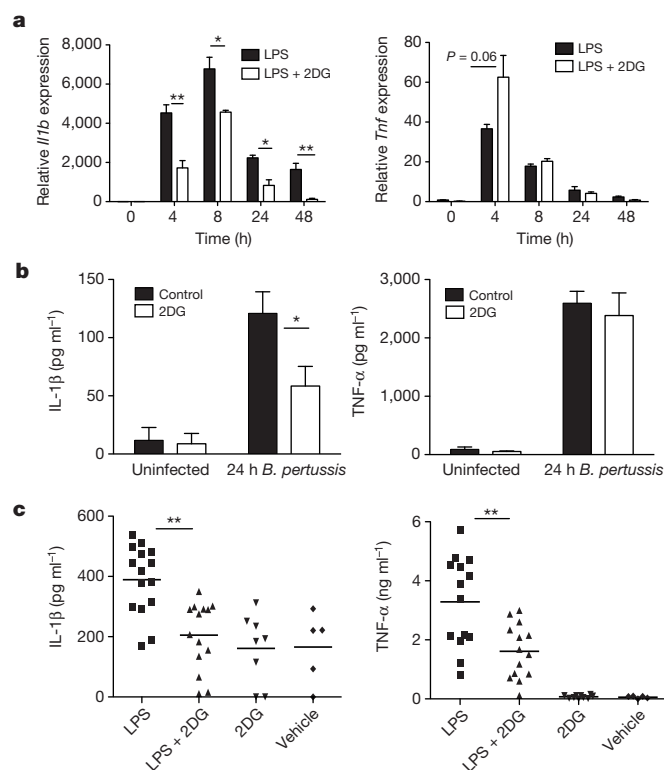


Figure 1 | Glycolysis is necessary for LPS-induced IL-1 β expression. a, b, *Il1b* and *Tnf* messenger RNA in 100 ng ml⁻¹ LPS- or 1 × 10⁵ *B. pertussis*-stimulated BMDMs ± 2DG (1 mM) pre-treatment for 3 h (*n* = 3). c, Mice injected intraperitoneally with or without 2DG (2 g kg⁻¹) or PBS for 3 h, then 15 mg kg⁻¹ LPS or PBS for 1.5 h. Serum concentrations of IL-1 β and TNF- α : LPS, *n* = 15; LPS + 2DG, *n* = 14; 2DG, *n* = 8; vehicle, *n* = 5. Error bars, s.e.m. **P* < 0.05; ***P* < 0.01.

¹School of Biochemistry and Immunology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin 2, Ireland. ²Department of Biological Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, USA. ³Centre for Computational and Integrative Biology, Massachusetts General Hospital, Richard B. Simches Research Center, Boston, Massachusetts 02114, USA. ⁴Apoptosis and Tumour Physiology Laboratory, The Beatson Institute for Cancer Research, Bearsden, Glasgow G61 1BD, UK. ⁵Medical Research Council Cancer Cell Unit Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 0X2, UK. ⁶The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁷Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA. ⁸Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. ⁹Academic Unit of Respiratory Medicine, Department of Infection and Immunity, University of Sheffield, Sheffield S10 2RX, UK. ¹⁰V. Nizet Laboratory, Division of Pediatrics, Centre for Neural Circuits and Behaviour, University of California, La Jolla, California 92093-0687, USA. ¹¹Conway Institute, University College Dublin, Dublin 4, Ireland. ¹²Inflammation Division, Walter and Eliza Hall Institute, 1G Royal Parade, Parkville, Victoria 3052, Australia.

*These authors contributed equally to this work.

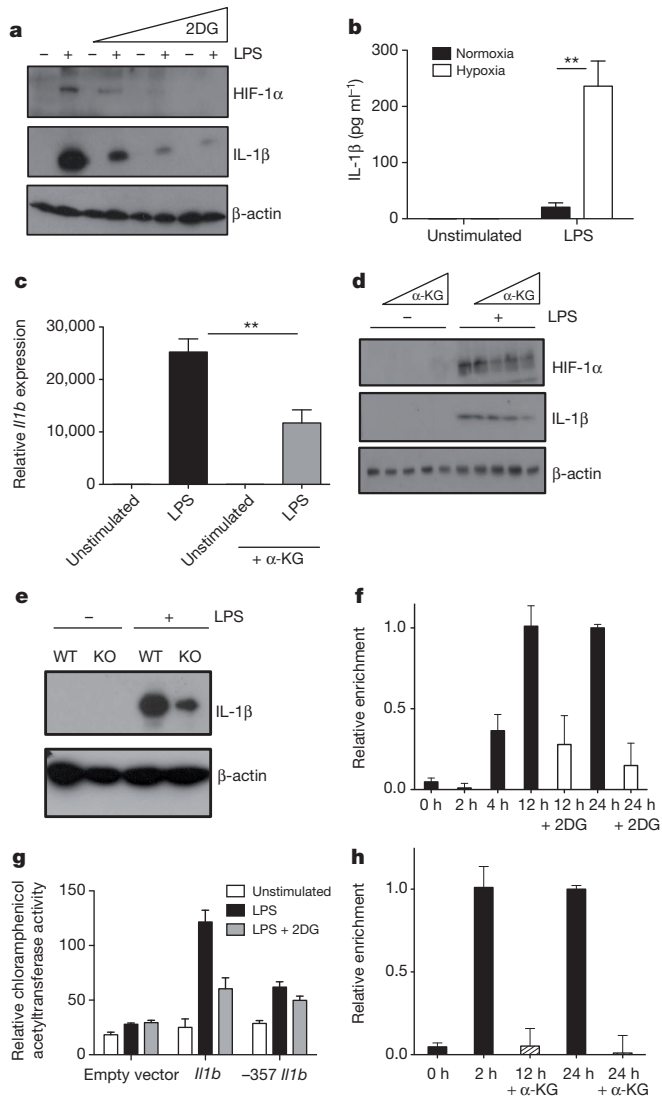


Figure 2 | HIF-1 α is responsible for LPS-induced IL-1 β expression. **a**, LPS-induced HIF-1 α and IL-1 β protein expression with or without 2DG (1, 5, 10 mM). **b**, IL-1 β in BMDMs incubated in normoxia or hypoxia for 24 h then LPS for 24 h. **c**, *Il1b* mRNA in LPS-stimulated BMDMs pre-treated with or without α -KG derivative³⁰ (1 mM). **d**, LPS-induced HIF-1 α and IL-1 β in BMDMs pre-treated with or without α -KG (0.01, 1, 100, 1000 μ M). **e**, LPS-induced IL-1 β protein in wild-type (WT) and *Hif1a* deficient (knockout, KO) BMDMs. **f**, ChIP-PCR using HIF-1 α antibody and primers specific for \sim 300 position of *Il1b* in LPS-treated BMDMs with or without 2DG (1 mM). **g**, Reporter activity in RAW-264 cells transfected with *Il1b* or \sim 357 *Il1b*. Representative of three experiments. Error bars, s.d. **h**, ChIP-PCR as above in BMDMs treated with LPS with or without α -KG (1 mM). Error bars, s.e.m. $^{**}P < 0.01$.

(Fig. 2d). Induction of IL-1 β was attenuated in *Hif1a* deficient macrophages (Fig. 2e).

Inspection of human (*IL1B*) and murine (*Il1b*) gene sequences showed a conserved canonical HIF-1 α binding site approximately 300 base pairs upstream from the transcription start site in the human gene, and at position \sim 357 in the mouse gene^{4,5}. Chromatin immunoprecipitation (ChIP)-PCR analysis in LPS-stimulated BMDMs showed HIF-1 α bound near the \sim 300 position of *Il1b* at 4, 12 and 24 h, which was inhibited by 2DG (Fig. 2f). LPS-induced *Il1b* luciferase activity, which was blocked by 2DG, had substantially reduced activity when \sim 357 in the *Il1b* (Fig. 2g) or \sim 300 in *IL1B* (Supplementary Fig. 9) promoter was mutated. LPS-induced HIF-1 α binding to the \sim 300 position of the *Il1b* promoter by ChIP analysis was abolished by

pre-treatment with α -KG (Fig. 2h). Therefore IL-1 β is a direct target of HIF-1 α , with HIF-1 α probably having a late effect more associated with the extended expression of IL-1 β given the kinetics, supporting previous data^{6,7}. The inhibition of IL-1 β but not TNF- α induction by 2DG is therefore explained by the HIF-1 α dependency in the *IL1B* gene. *Hif1a* deficiency also rescues mice from LPS-induced sepsis⁷ but how HIF-1 α protein is regulated by LPS is still unknown.

Several groups have shown stabilization of HIF-1 α by reactive oxygen species after LPS stimulation^{8,9}, which we confirmed (Supplementary Fig. 10). Also, HIF-1 α is stabilized by the PLC/PKC pathway⁹; however, treatment of BMDMs with specific inhibitors to PLC/PKC had no effect on LPS-induced HIF-1 α protein expression at 24 h (Supplementary Fig. 11).

Because both 2DG and α -KG could inhibit HIF-1 α accumulation and consequently induction of IL-1 β , we considered that the reported change in metabolism induced by LPS must be required for this response. We therefore next examined the metabolic profile of LPS-stimulated BMDMs by flux analysis, a metabolomic screen and microarray analysis. Extracellular flux analysis showed increased glucose use by LPS-stimulated BMDMs (Fig. 3a). This is due to increased glycolysis as measured by an increase in extracellular acidification rate accompanied by a decrease in oxygen consumption rate after LPS stimulation (Fig. 3a and Supplementary Fig. 12), confirming LPS induces the 'Warburg effect' of aerobic glycolysis.

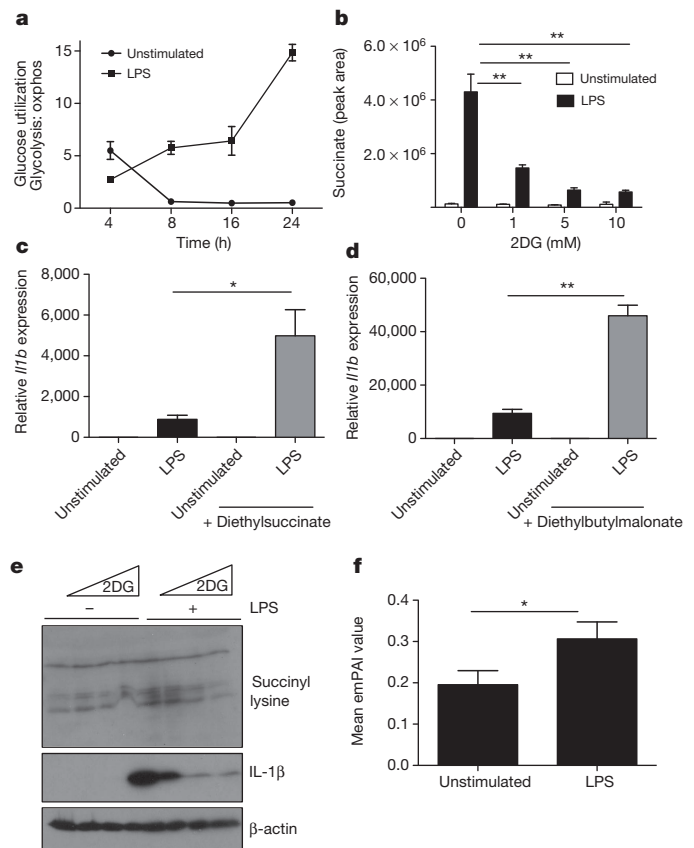


Figure 3 | Succinate is induced by LPS to drive HIF-1 α -induced IL-1 β expression. **a**, Glucose use over time, as a ratio of extracellular acidification rate/oxygen consumption rate in LPS-treated BMDMs, analysed on the Seahorse XF-24. **b**, Succinate abundance in LPS-stimulated BMDMs pre-treated with or without 2DG (1, 5, 10 mM). **c**, **d**, LPS-stimulated BMDMs pre-treated with or without 5 mM diethylsuccinate or diethylbutylmalonate (1 mM) for 3 h ($n = 3$). **e**, Succinyl-lysine expression and corresponding IL-1 β in LPS-stimulated BMDMs with or without 2DG (1, 5, 10 mM) pre-treatment for 3 h. **f**, Mean empAI values as a measure of relative abundance of proteins listed in Supplementary Table 1. Error bars, s.e.m. $^{*}P < 0.05$; $^{**}P < 0.01$.

The metabolomic screen confirmed this switch in metabolism with 73 metabolites changing out of 208 analysed (Supplementary Fig. 13). Glycolytic intermediates accumulated in 24 h of LPS stimulation. Despite decreased mitochondrial respiration, the tricarboxylic acid (TCA) cycle intermediates fumarate, malate and succinate accumulated. Succinate continued to accumulate between 4 and 24 h, and isocitrate was significantly depleted. Despite a minor increase in citrate at 24 h, there was significant accumulation of di- and triacylglycerols, indicating an increased rate of fatty-acid synthesis. Pentose phosphate pathway intermediates increased, as did the downstream products of purine and pyrimidine metabolism, guanosine, hypoxanthine and inosine (Supplementary Fig. 13). These substrates are necessary to meet the increased demand for energy and nucleic acid synthesis of these highly active cells.

Microarray analysis showed the metabolomic changes correlated with the gene expression profiles of LPS-stimulated BMDMs. Thirty-one metabolic enzyme/transporter-related genes were differentially expressed after 24 h of LPS treatment (Supplementary Fig. 14). The glucose transporters GLUT1, hexokinase 3, 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3, phosphoglucosyltransferase-2 and enolase 2 were significantly upregulated, confirming an increase in glycolysis. Lack of mitochondrial activity can be explained by decreased expression of malate dehydrogenase and isocitrate dehydrogenase. In addition, increased amounts of citrate and fatty acids suggested that activity was being diverted away from the TCA cycle for biosynthetic needs. The metabolic screen and flux analysis therefore showed that although there was an overall decrease in TCA cycle activity and mitochondrial respiration at 24 h LPS treatment, there was an accumulation of the key TCA cycle intermediate, succinate.

Further analysis by liquid chromatography–mass spectrometry (LC–MS) showed that LPS caused a 30-fold increase in succinate at 24 h, which equates to an increase in basal succinate from 25 ng to 1 µg per 2×10^6 cells. Pre-treatment with 2DG reduced intracellular succinate, in a dose-dependent manner (Fig. 3b).

Succinate can be transported from the mitochondria through the dicarboxylic acid transporter to the cytosol, where in excess it impairs PHD activity (by product inhibition) leading to HIF-1 α stabilization and activation. This phenomenon has been defined as ‘pseudohypoxia’¹⁰. This effect is blocked by α -KG, the substrate for PHD that generates succinate as a by-product in HIF-1 α hydroxylation^{10,11}. We had observed that α -KG inhibited stabilization of HIF-1 α and induction of IL-1 β (Fig. 2), indicating that the stabilization of HIF-1 α by LPS was probably dependent on the rise of succinate. We tested this further and found that cell-permeable diethylsuccinate alone had no effect on expression of HIF-1 α (Supplementary Fig. 15) or *Il1b* (Fig. 3c), because LPS is required to induce *Hif1a* mRNA (Supplementary Fig. 6). However, in combination with LPS, succinate profoundly augmented *Il1b* mRNA expression (Fig. 3c). This supports previous work in which succinate boosted LPS-induced IL-1 β in dendritic cells¹². Succinate amounts are regulated by succinate dehydrogenase. Diethylbutylmalonate, a succinate dehydrogenase inhibitor, increased LPS-induced *Il1b* mRNA (Fig. 3d). Both diethylsuccinate and diethylbutylmalonate increased HIF-1 α and IL-1 β protein in LPS-treated BMDMs (Supplementary Fig. 15). Diethylsuccinate also increased mRNA of the LPS-induced HIF-1 α target *Phd3* (also known as *Egln3*) (ref. 13) (Supplementary Fig. 16). The effect of diethylsuccinate on IL-1 β was due to HIF-1 α as LPS-stimulated BMDMs from *Hif1a* deficient mice had less IL-1 β protein and mRNA than wild type, and pre-treatment with diethylsuccinate had no significant effect. IL-6 and TNF- α were unaffected in *Hif1a* deficient macrophages (Supplementary Fig. 17). Therefore LPS-induced succinate can act as a signal to increase IL-1 β expression through HIF-1 α .

Certain metabolic enzymes have been shown to undergo succinylation¹⁴. We observed an increase in succinylation of proteins in response to LPS, which was reduced by 2DG in a dose-dependent manner (Fig. 3e). This was confirmed by a [³²P]NAD assay (Supplementary

Fig. 18). LPS treatment resulted in a twofold increase in protein succinylation, which was inhibited by 36% with 2DG treatment (data not shown). The expression of *Sirt5* (a recently reported desuccinylase and demalonylase¹⁴) was inhibited by LPS, suggesting reduced desuccinylase activity (Supplementary Fig. 19). BMDMs treated for 24 h with LPS also had a significantly altered NAD/NADH ratio favouring NADH, thus providing little NAD⁺ substrate for SIRT5 activity (Supplementary Fig. 20). This decrease in the NAD/NADH ratio also provides further evidence of enhanced glycolysis and lower respiration. The increase in protein succinylation could therefore be explained by an increase in succinate and a decrease in the expression and activity of the desuccinylase SIRT5.

The relative abundance of proteins in BMDM lysates enriched using a succinyl lysine antibody were detected using LC–MS. Figure 3f represents the mean exponentially modified protein abundance index (emPAI) values (relative quantity) of proteins from unstimulated and LPS-stimulated samples, an enrichment of succinylated proteins being evident in the LPS-treated BMDMs. The increased succinylation of malate dehydrogenase (Supplementary Table 1) has previously been reported in animal tissue¹⁴. Other proteins identified include glyceraldehyde-3-phosphate dehydrogenase, glutamate carrier 1, L-lactate dehydrogenase A chain and transaldolase. The consequence of succinylation of these proteins, a covalent modification activated by LPS not previously reported, is now under investigation.

A dysfunctional TCA cycle in LPS-stimulated BMDMs pointed towards an alternative source of succinate. The microarray screen showed significantly higher concentrations of glutamine transporter (*Slc3a2*) at 24 h. Succinate can be derived from glutamine through anaplerosis by α -KG. We confirmed LPS induction of *Slc3a2* mRNA (Fig. 4a). *SLC3A2* mRNA expression has previously been shown to increase during intestinal inflammation and is regulated by Sp1 and NF- κ B¹⁵. Blocking the NF- κ B pathway with a specific MyD88 inhibitory peptide significantly reduced LPS-induced *Slc3a2* mRNA expression (Supplementary Fig. 21). Knockdown of *SLC3A2* by short interfering RNA in human PBMCs (Supplementary Fig. 22) significantly reduced LPS-induced IL-1 β expression (Fig. 4b), but had no effect on IL-6 or TNF- α (Supplementary Fig. 22). This was also demonstrated in the RAW-264 macrophage cell line (Supplementary Fig. 23).

In addition to anaplerosis, succinate can also be derived from glutamine through the ‘GABA (γ -aminobutyric acid) shunt’ pathway. Gene expression analysis and metabolomics showed LPS-induced GABA transporters (*SLC6A13*, *SLC6A12*) and increased GABA production (Supplementary Fig. 13). The increase in GABA was further confirmed by LC–MS (Fig. 4c).

Next we traced the metabolism of [¹³C₅,¹⁵N₂]glutamine and measured the increase in [¹³C₄]succinate abundance in BMDMs treated with LPS for 24 h (Fig. 4d). This showed a marked increase in succinate production from labelled glutamine. Of the total succinate pool in the LPS-treated BMDMs, $18.4 \pm 0.9\%$ ($n = 4$; \pm SEM) was labelled after 3 h incubation, as shown by the ratio of [¹³C₄]succinate (heavy labelled succinate) to the sum of all measured succinate isotopomers. This represents a 14-fold increase in succinate relative to non-LPS-treated BMDMs. Treatment of BMDMs with the specific, irreversible inhibitor of the key ‘GABA shunt’ enzyme GABA transaminase, vigabatrin¹⁶, reduced the percentage of labelled succinate to $16.0 \pm 0.6\%$ ($n = 4$; $P = <0.05$), representing a 9.4-fold elevation over the concentrations of succinate attained with vigabatrin alone. This represented a 33% decrease compared with the samples from BMDMs not treated with vigabatrin (Fig. 4e). The elevation in succinate in response to LPS derived from glutamine is therefore largely from anaplerosis by α -KG although a proportion is derived from the ‘GABA shunt’.

Pre-treatment of BMDMs with vigabatrin significantly reduced LPS-induced HIF-1 α and IL-1 β protein (Fig. 4f) as well as *Phd3* mRNA (Supplementary Fig. 24). The inhibitory effect was specific for IL-1 β as vigabatrin inhibited *Il1b* mRNA expression (Fig. 4g) but not TNF- α protein amounts or *IL6* mRNA (Supplementary Fig. 24).

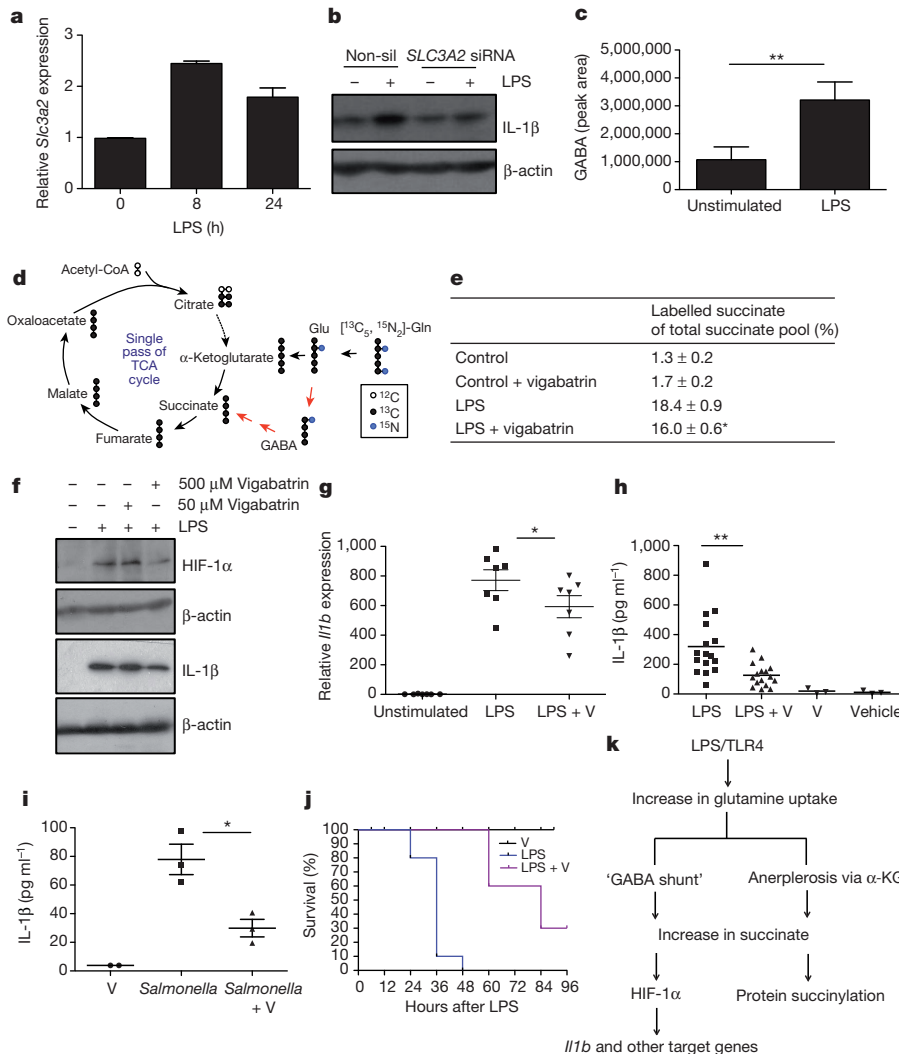


Figure 4 | Glutamine is the source of LPS-induced succinate. **a**, *Scl3a2* mRNA in LPS-treated BMDMs for 8 and 24 h. **b**, IL-1β in human PBMCs transfected with 100 nM *SLC3A2* short interfering RNA compared with non-silencing (non-sil.) control. Representative of three experiments. **c**, GABA abundance in serum-deprived BMDMs stimulated with 10 ng ml⁻¹ LPS for 24 h. **d**, Tracing of [¹³C₅, ¹⁵N₂]glutamine in BMDMs treated with LPS for 24 h. **e**, LPS-stimulated BMDMs pre-treated with or without vigabatrin (500 μM) for 20 h then 1 mM [¹³C₅, ¹⁵N₂]glutamine for 4 h. Table represents ratio of [¹³C₄]succinate to the sum of all measured succinate isotopomers as measured by LC-MS. Values are mean ± s.e.m. (*n* = 4). **f**, **g**, LPS-induced HIF-1α and IL-1β protein and *Il1b* mRNA (*n* = 7) in BMDMs pre-treated with or without vigabatrin for 30 min. **h**, IL-1β in serum from mice injected intraperitoneally

with or without vigabatrin (400 mg kg⁻¹) or PBS for 1.5 h, then 15 mg kg⁻¹ LPS or PBS solution for 1.5 h (LPS, *n* = 16; LPS + vigabatrin (LPS + V), *n* = 14; vigabatrin (V), *n* = 3; vehicle, *n* = 3). **i**, IL-1β in serum from mice injected intraperitoneally with or without vigabatrin (400 mg kg⁻¹) or PBS for 1.5 h then infected with 1 × 10⁶ *Salmonella typhimurium* UK1 intraperitoneally for 2 h. **j**, Survival of mice injected intraperitoneally with or without vigabatrin (400 mg kg⁻¹) or PBS for 1.5 h, then 60 mg kg⁻¹ LPS or PBS (PBS, *n* = 10 (not shown); vigabatrin, *n* = 10; LPS, *n* = 10; LPS + vigabatrin, *n* = 10). Error bars, s.e.m. **P* < 0.05; ***P* < 0.01. **k**, Proposed model: LPS induces high concentrations of succinate to drive HIF-1α-dependent *Il1b* expression and protein succinylation.

This was also evident *in vivo* as a prescription formulation of vigabatrin (Sabril) reduced LPS-induced IL-1β (Fig. 4h) but had no effect on IL-6 (Supplementary Fig. 25). Similar to 2DG, the inhibitory effect of vigabatrin on TNF-α is probably due to downstream signalling of IL-1β (Supplementary Fig. 25). Similarly vigabatrin reduced IL-1β serum amounts in response to *Salmonella* infection (Fig. 4i). This was consistent with the greater bacterial loading in the spleens of vigabatrin-treated mice (Supplementary Fig. 26). In a model of sepsis, vigabatrin also protected mice, presumably by blocking HIF-1α activation (*P* < 0.001) (Fig. 4j).

Our study therefore shows that chronic activation of macrophages with LPS causes an increase in intracellular succinate by glutamine-dependent anerplerosis and the 'GABA-shunt' pathway. Succinate acts as an endogenous danger signal to stabilize HIF-1α, which in turn specifically regulates gene expression of *Il1b* and other HIF-1α-dependent

genes and leads to protein succinylation (Fig. 4k). The role of the 'GABA shunt' is also of substantial interest. Vigabatrin is used clinically to treat epilepsy, as it can increase GABA, which is an inhibitory neurotransmitter. However, adverse effects include significant increased risk of infections¹⁷, which could be caused by limitation of the GABA shunt in macrophages.

Succinate therefore joins other signals derived from mitochondria, including cytochrome C (ref. 18), and mitochondrial DNA¹⁹ that have a role in signalling cell trauma. Succinate can be released to signal through GPR91, which synergizes with Toll-like receptors¹². High succinate concentrations have been detected in the plasma of patients with peritonitis²⁰, and in the urine and plasma of diabetic and metabolic disease rodent models^{21,22}. In addition, patients harbouring mutations in succinate dehydrogenase have increased HIF-1α activity^{23,24} and circulating succinate²⁵. Succinate activates HIF-1α in tumours^{10,26},

indicating an important similarity between inflammation and cancer. The inflammatory process may have a tumorigenic effect by increasing succinate. The identification of succinate as an inflammatory signal may therefore be important for our understanding of innate immunity in both inflammatory diseases and cancer.

METHODS SUMMARY

Female C57BL/6 mice were from Harlan UK. BMDMs from *Hif1a^{flox/flox}/LysMcre* and wild-type controls, *Hif1a^{flox/flox}* littermate controls, were obtained from M. Whyte. All experiments had previous ethical approval from Trinity College Dublin Animal Research Ethics Committee. BMDMs were prepared as described previously²⁷. Unless stated, 100 ng ml⁻¹ LPS and 5 × 10⁵ cells per millilitre BMDMs were used in *in vitro* experiments. Two-tailed *t*-tests measured significance of expression of genes with different treatments. Western blot analysis was performed as previously described²⁸. Reactive oxygen species were measured on LPS-stimulated BMDMs as previously described²⁹. ELISAs for IL-1β (R&D, Quantikine), IL-6 and TNF-α (R&D, DuoSet) were performed according to the manufacturer's instructions. Microarray was performed on the Illumina microarray platform. Genes with fold changes in expression greater than 1.5 and Benjamini–Hochberg adjusted *t*-test-based *P* values less than 0.05 were considered significant. Metabolic profiling was in the positive ion MS mode using an LC system. Metabolites with fold change in amounts greater than 10% and *P* values less than 0.05 were considered to be statistically significant. Succinate and GABA amounts were measured by LC–MS analysis. ChIP–PCR using HIF-1α antibody and primers specific for the –300 position of the *Il1b* and *IL1B* genes were performed using a modification of the Millipore/Upstate protocol (MCPROTO407). PBMCs were isolated and transfected with short interfering RNA as described previously²⁸ using Hs_SLC3A2_1, target sequence AATCCTGAGCCTACTCGAATC (Qiagen). RAW-264 cells were transfected as described previously²⁸ using Mm_Slc3a2_1 target sequence CAGAAGGATGAAATCAATGAA (Qiagen). The plasmid containing human *IL1B*, pGL3-*IL1* (a gift from J. Kim) and a chloramphenicol acetyltransferase plasmid containing –4093 to +45 of the mouse *Il1b* were mutated using the QuikChange lightning site directed mutagenesis kit (Agilent) according to the manufacturer's instructions. The NAD/NADH ratio was analysed according to the manufacturer's instructions (Biovision K337-100).

Full Methods and any associated references are available in the online version of the paper.

Received 28 March 2012; accepted 5 February 2013.

Published online 24 March 2013.

- Rodriguez-Prados, J. C. *et al.* Substrate fate in activated macrophages: a comparison between innate, classic, and alternative activation. *J. Immunol.* **185**, 605–614 (2010).
- Krawczyk, C. M. *et al.* Toll-like receptor-induced changes in glycolytic metabolism regulate dendritic cell activation. *Blood* **115**, 4742–4749 (2010).
- Pan, H. & Wu, X. Hypoxia attenuates inflammatory mediators production induced by *Acanthamoeba* via Toll-like receptor 4 signaling in human corneal epithelial cells. *Biochem. Biophys. Res. Commun.* **420**, 685–691 (2012).
- Hiscott, J. *et al.* Characterization of a functional NF-κB site in the human interleukin 1β promoter: evidence for a positive autoregulatory loop. *Mol. Cell. Biol.* **13**, 6231–6240 (1993).
- Ghisletti, S. *et al.* Parallel SUMOylation-dependent pathways mediate gene- and signal-specific transrepression by LXRs and PPARγ. *Mol. Cell* **25**, 57–70 (2007).
- Zhang, W. *et al.* Evidence that hypoxia-inducible factor-1 (HIF-1) mediates transcriptional activation of interleukin-1β (IL-1β) in astrocyte cultures. *J. Neuroimmunol.* **174**, 63–73 (2006).
- Peyssonnaud, C. *et al.* Cutting edge: essential role of hypoxia inducible factor-1α in development of lipopolysaccharide-induced sepsis. *J. Immunol.* **178**, 7516–7519 (2007).
- Chandel, N. S. *et al.* Reactive oxygen species generated at mitochondrial complex III stabilize hypoxia-inducible factor-1α during hypoxia: a mechanism of O₂ sensing. *J. Biol. Chem.* **275**, 25130–25138 (2000).
- Sumbayev, V. V. LPS-induced Toll-like receptor 4 signalling triggers cross-talk of apoptosis signal-regulating kinase 1 (ASK1) and HIF-1α protein. *FEBS Lett.* **582**, 319–326 (2008).
- Selak, M. A. *et al.* Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-α prolyl hydroxylase. *Cancer Cell* **7**, 77–85 (2005).
- Koivunen, P. *et al.* Inhibition of hypoxia-inducible factor (HIF) hydroxylases by citric acid cycle intermediates: possible links between cell metabolism and stabilization of HIF. *J. Biol. Chem.* **282**, 4524–4532 (2007).
- Rubic, T. *et al.* Triggering the succinate receptor GPR91 on dendritic cells enhances immunity. *Nature Immunol.* **9**, 1261–1269 (2008).
- Walmsley, S. R. *et al.* Prolyl hydroxylase 3 (PHD3) is essential for hypoxic regulation of neutrophilic inflammation in humans and mice. *J. Clin. Invest.* **121**, 1053–1063 (2011).
- Du, J. *et al.* Sirt5 is a NAD-dependent protein lysine demalonylase and desuccinylase. *Science* **334**, 806–809 (2011).
- Yan, Y., Dalmasso, G., Sitaraman, S. & Merlin, D. Characterization of the human intestinal CD98 promoter and its regulation by interferon-γ. *Am. J. Physiol. Gastrointest. Liver Physiol.* **292**, G535–G545 (2007).
- Choi, S. & Silverman, R. B. Inactivation and inhibition of γ-aminobutyric acid aminotransferase by conformationally restricted vigabatrin analogues. *J. Med. Chem.* **45**, 4531–4539 (2002).
- Walker, S. D. & Kalvainen, R. Non-vision adverse events with vigabatrin therapy. *Acta Neurol. Scand. (suppl.)*, **124**, 72–82 (2011).
- Bossy-Wetzel, E., Newmeyer, D. D. & Green, D. R. Mitochondrial cytochrome c release in apoptosis occurs upstream of DEVD-specific caspase activation and independently of mitochondrial transmembrane depolarization. *EMBO J.* **17**, 37–49 (1998).
- Zhang, Q. *et al.* Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature* **464**, 104–107 (2010).
- Rotstein, O. D., Pruett, T. L. & Simmons, R. L. Lethal microbial synergism in intra-abdominal infections. *Escherichia coli* and *Bacteroides fragilis*. *Arch. Surg.* **120**, 146–151 (1985).
- Toma, I. *et al.* Succinate receptor GPR91 provides a direct link between high glucose levels and renin release in murine and rabbit kidney. *J. Clin. Invest.* **118**, 2526–2534 (2008).
- Sadagopan, N. *et al.* Circulating succinate is elevated in rodent models of hypertension and metabolic disease. *Am. J. Hypertens.* **20**, 1209–1215 (2007).
- Gimenez-Roqueplo, A. P. *et al.* The R22X mutation of the *SDHD* gene in hereditary paraganglioma abolishes the enzymatic activity of complex II in the mitochondrial respiratory chain and activates the hypoxia pathway. *Am. J. Hum. Genet.* **69**, 1186–1197 (2001).
- Dahia, P. L. *et al.* A HIF1α regulatory loop links hypoxia and mitochondrial signals in pheochromocytomas. *PLoS Genet.* **1**, 72–80 (2005).
- Hobert, J. A., Mester, J. L., Moline, J. & Eng, C. Elevated plasma succinate in PTEN, SDHB, and SDHD mutation-positive individuals. *Genet. Med.* **14**, 616–619 (2012).
- Pistollato, F. *et al.* Hypoxia and succinate antagonize 2-deoxyglucose effects on glioblastoma. *Biochem. Pharmacol.* **80**, 1517–1527 (2010).
- Masters, S. L. *et al.* Activation of the NLRP3 inflammasome by islet amyloid polypeptide provides a mechanism for enhanced IL-1β in type 2 diabetes. *Nature Immunol.* **11**, 897–904 (2010).
- Doyle, S. L. *et al.* The GOLD domain-containing protein TMED7 inhibits TLR4 signalling from the endosome upon LPS stimulation. *Nature Commun.* **3**, 707 (2012).
- West, A. P. *et al.* TLR signalling augments macrophage bactericidal activity through mitochondrial ROS. *Nature* **472**, 476–480 (2011).
- MacKenzie, E. D. *et al.* Cell-permeating α-ketoglutarate derivatives alleviate pseudohypoxia in succinate dehydrogenase-deficient cells. *Mol. Cell. Biol.* **27**, 3282–3289 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the European Research Council, Science Foundation Ireland, the Health Research Board, European Union FP7 programme 'TIMER', Wellcome Trust, National Institutes of Health, Helmsley Trust, Nestle Research Centre, VESKI, The Duquesne University Hunkele Dreded Disease Award, The Interleukin Foundation and the National Health and Medical Research Council for funding. We also thank R. Thompson for assistance with *Hif1a*^{-/-} mice and M. Murphy for discussions.

Author Contributions G.M.T. designed and did experiments, analysed data and wrote the paper; L.A.J.O. conceived ideas and oversaw the research programme; A.M.C., E.M.P., A.F.M. and J.A. designed and did experiments and analysed data; C.F., N.J.B., B.K., N.H.F., L.Z., A.G., Z.T., S.S.J., S.C.C., S.W., K.P. and F.C.B. did experiments; G.G., R.J.X., C.C., M.H. and B.E.C. performed bioinformatic analysis; E.C., V.N., M.W., C.T.T., H.L., S.L.M., E.G., V.K. and C.C., provided advice and reagents; P.E.A. and R.J.X. conceived ideas and oversaw a portion of the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.A.J.O. (laoneill@tcd.ie).

METHODS

Reagents. LPS used in *in vitro* and *in vivo* studies was from *Escherichia coli*, serotype EH100 (Alexis) and 055:B5 (Sigma-Aldrich), respectively. 2DG (D3179), diethyl succinate (112402), diethyl butylmalonate (112402), vigabatrin (S(+)- γ -vigabatrin, V113), Gö6983 (PKC inhibitor) and *N*-acetyl cysteine (NAC) were all purchased from Sigma, U73122 (PLC inhibitor) from Calbiochem, dimethylallyl glycine (71210) from Cayman Chemicals; α -KG was a gift from E. MacKenzie, and Sabril was a gift from L. Cleary. Antibodies used were anti-succinyl lysine (PTM Biolab, PTM-401), anti-IL-1 β (R&D, AF401-NA) and anti-HIF-1 α (Novus, NB100-449).

Microarray profiling. Quantification of RNA concentration and purity was measured using a NanoDrop spectrophotometer (Thermo Scientific). The quality of the mouse RNA was ascertained using an Agilent Bioanalyser 2100 with the NanoChip protocol. A total of 500 ng RNA was amplified and labelled using the Illumina TotalPrep RNA Amplification kit (Ambion) according to the manufacturer's instructions. A total of 1.5 μ g of labelled cRNA was then prepared for hybridization to the Illumina Mouse WG-6 chip. The Illumina microarray was performed by Partners HealthCare Center for Personalized Genetic Medicine Core.

Cellular respiration and extracellular acidification. An XF24 Extracellular Flux analyser (Seahorse Biosciences) was used to determine the bioenergetic profile of LPS-stimulated BMDMs. BMDMs were plated at 200,000 cells per well in XF24 plates overnight before being stimulated for up to 24 h with LPS. Oxygen consumption rate and extracellular acidification rate were assessed in glucose-containing media (Seahorse Biosciences). Results were normalized to cell number.

Measurement of metabolites by LC-MS. Endogenous metabolic profiling was performed in BMDMs stimulated with LPS for 4 and 24 h with or without 2DG. Polar metabolites were extracted in 80% methanol. Metabolic profiles were obtained using three LC-MS methods. Two separate hydrophilic interaction liquid chromatography (HILIC) methods were used to measure polar metabolites in positive and negative ionization mode MS, and one reversed phase method was used to profile lipids in the positive ion mode. Polar metabolites were profiled in the positive ion MS mode using an LC system comprising a 1200 Series Pump (Agilent Technologies) and an HTS PAL autosampler (Leap Technologies) that was coupled to a 4000 QTRAP mass spectrometer (AB SCIEX) equipped with an electrospray ionization source. Samples were prepared by drying 100 μ l of cell extracts under nitrogen and re-suspending the residue in 100 μ l of 10/67.4/22.4/0.2 v/v/v/v water/acetonitrile/methanol/formic acid containing stable-isotope-labelled internal standards (valine-d8, Sigma-Aldrich; and phenylalanine-d8, Cambridge Isotope Laboratories). The samples were centrifuged (10 min, 1610g, 4 °C), and the supernatants were injected directly onto a 150 mm \times 2.1 mm Atlantis HILIC column (Waters). The column was eluted isocratically with 5% mobile phase A (10 mM ammonium formate/0.1% formic acid) for 1 min followed by a linear gradient to 40% mobile phase B (acetonitrile/0.1% formic acid) over 10 min. The ion spray voltage was 4.5 kV and the source temperature was 425 °C. All metabolites were measured using several reaction monitoring scans (MRM). MS settings, including declustering potentials and collision energies, for each metabolite were optimized by infusion of reference standards before sample analyses.

Metabolites labelled with [$^{13}\text{C}_5$, $^{15}\text{N}_2$]glutamine were analysed using negative ion mode MS performed using an ACQUITY UPLC (Waters) coupled to a 5500 QTRAP triple quadrupole mass spectrometer (AB SCIEX). Cell extracts in 80% methanol were directly injected onto a 150 mm \times 2.0 mm Luna NH2 column (Phenomenex) that was eluted at a flow rate of 400 μ l min $^{-1}$ with initial conditions of 10% mobile phase A (20 mM ammonium acetate and 20 mM ammonium hydroxide in water) and 90% mobile phase B (10 mM ammonium hydroxide in 75:25 v/v acetonitrile/methanol) followed by a 10 min linear gradient to 100%

mobile phase A. The ion spray voltage was -4.5 kV and the source temperature was 500 °C. All data were acquired using MRM scanning. Lipids were profiled in the positive ion mode using an 1100 Series pump and autosampler (Agilent Technologies) coupled to a QSTAR-XL MS system (AB SCIEX).

MultiQuant software (version 1.2; AB SCIEX) was used to process all raw LC-MS data and integrate chromatographic peaks. The processed data were manually reviewed for quality of integration and compared against known standards to confirm metabolite identities.

Succinate and GABA measurements: after stimulations, BMDMs were lysed with a solution kept in dry ice/methanol (-80 °C) composed of 50% methanol and 30% acetonitrile in water and quickly scraped. The insoluble material was immediately pelleted in a cooled centrifuge (0 °C) and the supernatant collected for subsequent LC-MS analysis. A ZIC-HILIC column (4.6 mm \times 150 mm, guard column 2.1 mm \times 20 mm, Merck) was used for LC separation using formic acid, water acetonitrile as component of the mobile phase.

[^{32}P]NAD assay. BMDMs were lysed for 30 min at 4 °C in ice-cold lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA, 0.1% NP-40, 10% glycerol), spun at 1700g for 10 min and the supernatant was collected for trypsin digestion. Protein (300 μ g) from each sample was dissolved in 6 M guanidine-HCl, 50 mM Tris-HCl (pH 8.0), 15 mM DTT in a reaction volume of 450 μ l. Iodoacetamide (1 M, 22.5 μ l) was added and the mixture was incubated at room temperature with gentle shaking for 1 h. DTT (1 M, 20 μ l) was added with gentle mixing for 1 h. Trypsin (10 μ g) (Promega, V51111) was added in the final buffer system (50 mM Tris pH 7.4, 1 mM CaCl $_2$) to digest proteins for 16 h at 37 °C. The reaction was quenched by adding 67.5 μ l 10% TFA and the digested peptides were desalted by Sep-Pack C18 cartridge (1 cm 3 50 mg $^{-1}$; Waters) and lyophilized. To detect the protein succinylation concentration in macrophages \pm LPS SIRT5 was used to hydrolyse succinyl peptides using [^{32}P]NAD. Reactions were performed in 10 μ l solutions with 1 μ Ci [^{32}P]NAD (ARC, ARP 0141, 800 Ci mmol $^{-1}$, 0.125 μ M), 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM DTT. 100 μ M synthetic H3K9 succinyl and malonyl peptides were used as positive controls. Macrophage trypsin-digested peptides (100 μ g) were used as substrates. The reactions were incubated with 4 μ M SIRT5 at 37 °C for 2 h. A total of 0.5 μ l of each reaction mixture were spotted onto silica gel thin-layer chromatography plates and developed with 7:3 ethanol: ammonium bicarbonate (1 M aqueous solution). The plates were further air-dried and exposed by PhosphorImaging screen (GE Healthcare) overnight. The signal was detected using a STORM860 phosphorimager (GE Healthcare).

Identification of proteins by LC-MS/MS in a succinyl-lysine pull-down. LPS-treated BMDMs were lysed as above. Anti-succinyl lysine antibody, or non-specific rabbit IgG in the control sample, were pre-coupled to protein A/G agarose beads, washed and incubated with the cell lysates for 2 h at 4 °C. The immune complexes were precipitated and washed thoroughly in lysis buffer. The protein was eluted by adding sample buffer and subsequently separated by SDS-polyacrylamide gel electrophoresis. Gel slices ranging from 35–40 kDa and 55–65 kDa were subjected to one-dimensional nano-scale LC-MS/MS mass spectrometry using an LTQ Orbitrap Velos Pro. Peptides were identified with Mascot (Matrix Sciences) against the Mouse IPI database with a protein score cut-off of 37. The emPAI values of proteins of appropriate molecular mass, excluding non-specifically bound proteins with highest values in the IgG control sample as well as keratin and immunoprecipitation antibody contaminations, were used to estimate relative abundance in the samples. Enrichment of proteins was statistically analysed using an unpaired, two-sided *t*-test for the mean emPAI values.

Endotoxin-induced model of sepsis. Mice were treated \pm vigabatrin (400 mg kg $^{-1}$) or PBS for 30 min. Sepsis was induced by injecting 60 mg kg $^{-1}$ of LPS and survival was monitored. Mice were killed immediately at a humane end-point.

Structural and energetic basis of folded-protein transport by the FimD usher

Sebastian Geibel^{1*}, Erik Procko^{2*}, Scott J. Hultgren³, David Baker² & Gabriel Waksman¹

Type 1 pili, produced by uropathogenic *Escherichia coli*, are multi-subunit fibres crucial in recognition of and adhesion to host tissues¹. During pilus biogenesis, subunits are recruited to an outer membrane assembly platform, the FimD usher, which catalyses their polymerization and mediates pilus secretion². The recent determination of the crystal structure of an initiation complex provided insight into the initiation step of pilus biogenesis resulting in pore activation, but very little is known about the elongation steps that follow³. Here, to address this question, we determine the structure of an elongation complex in which the tip complex assembly composed of FimC, FimF, FimG and FimH passes through FimD. This structure demonstrates the conformational changes required to prevent backsliding of the nascent pilus through the FimD pore and also reveals unexpected properties of the usher pore. We show that the circular binding interface between the pore lumen and the folded substrate participates in transport by defining a low-energy pathway along which the nascent pilus polymer is guided during secretion.

All pilus subunits exhibit an incomplete immunoglobulin-like fold lacking β -strand G, leaving a large hydrophobic groove across the subunit surface^{4–7}. Folding of pilus subunits is catalysed by a periplasmic chaperone⁸ (FimC for type 1 pili) (Supplementary Fig. 1a). In a mechanism known as donor-strand complementation, FimC donates the missing β -strand *in trans*, thereby complementing and stabilizing the pilus subunit's fold^{4,6} (Supplementary Fig. 1b). The resulting binary chaperone-subunit complexes are next recruited to an outer membrane assembly 'nanomachine', the usher (FimD for type 1 pili), for pilus assembly and secretion. Subunit polymerization occurs through a mechanism known as donor-strand exchange (DSE), whereby the amino-terminal extension (Nte) of the subunit next in assembly replaces the chaperone strand in the groove of the previously assembled (or receiving) subunit⁵ (Supplementary Fig. 1b).

Ushers consist of five functional domains: an N-terminal domain (NTD); a pore domain that, in the resting state, is occluded by a plug domain; and two carboxy-terminal domains (CTD1 and CTD2). Recently, the crystal structure of the pilus biogenesis initiation complex FimD–FimC–FimH (Supplementary Fig. 1c) inspired a model for subunit recruitment, polymerization and secretion by the usher³ (Supplementary Fig. 1d). FimH consists of two domains: a pilin domain (FimH_p) similar to all other pilus subunits, and a receptor-binding domain referred to as the lectin domain⁶ (FimH_L). In the FimD–FimC–FimH structure, FimH_L is located inside the FimD pore. FimH_p is still in donor-strand complementation interaction with the chaperone FimC bound to CTD1 and CTD2. The usher NTD, stabilized by the plug domain, is free and serves as a recruitment site for the next chaperone-subunit complex to be assembled (FimC–FimG; Supplementary Fig. 1d). Binding to the NTD places the FimG Nte peptide in an ideal position for DSE to occur³, leading to the dissociation of FimC bound to FimH_p. FimC has no binding affinity for the usher CTDs on its own, and so dissociates from the complex. At this

point, the entire nascent pilus translocates up inside the pore and the chaperone-subunit complex bound to the NTD transfers to the CTDs in a handover step (Supplementary Fig. 1d). The usher NTD is now free to enter another cycle of subunit incorporation by recruiting the next chaperone-subunit complex for assembly (FimC–FimF). This model for the subunit-incorporation cycle remains, however, to be structurally documented. Here we describe the structure of the elongation complex FimD–FimC–FimF–FimG–FimH, a post-initiation state during pilus biogenesis.

The complex FimD–FimC–FimF–FimG–FimH was purified to homogeneity and crystallized, and its structure was determined to a resolution of 3.8 Å (Supplementary Tables 1 and 2 and Supplementary Figs 2–4). The crystal structure of FimD–FimC–FimF–FimG–FimH captures the secretion of the type 1 pilus tip consisting of pilus subunits FimF, FimG and FimH through the usher pore across the outer bacterial membrane (Fig. 1a). The adhesin FimH is incorporated at the distal end of the pilus tip and found on the extracellular side of the transporter. FimG occupies the lumen of the usher β -barrel and contacts the usher CTD1, plug and β -barrel domains. As the last-assembled pilus subunit, FimF is found still in donor-strand complementation interaction with its chaperone FimC, which is bound by the usher C-terminal domains CTD1 and CTD2 in the periplasm. Thus, the structure provides the remarkable view of a transporter entirely traversed by its protein polymer substrate, with one subunit having entirely emerged from the transporter and another still remaining on the cellular side of the transporter. The type 1 pilus tip subunits FimH, FimG and FimF are linked non-covalently by DSE, where the N-terminal extension of FimF is inserted into FimG and that of FimG is inserted into FimH_p. The usher NTD shares only an interface with the plug domain and is otherwise exposed. Modelling shows that the NTD is indeed available to bind the next incoming chaperone-subunit complex, FimC–FimA, bringing it into close proximity to FimF and in the appropriate orientation to undergo DSE (Fig. 1b). This is consistent with the DSE model for pilus subunit incorporation introduced in ref. 3.

When the FimH proteins in the initiation complex FimD–FimC–FimH and the elongation complex FimD–FimC–FimF–FimG–FimH are superimposed, it becomes apparent that FimH undergoes a large conformational change when exiting the usher pore, with the angle between its two domains (FimH_p and FimH_L) decreasing⁹ (Fig. 2a and Supplementary Fig. 5). We speculate that such a conformational change may provide the necessary energy to favour translocation of FimH outside the pore and possibly prevent motion back into the periplasm. Another remarkable structural rearrangement affecting FimH during transport is the compression and decompression of the FimH_L domain structure before and after transport, respectively (Fig. 2b).

Using the physically realistic ROSETTA energy function^{10,11}, we next investigated the interactions between the usher pore and its inserted substrate, FimH_L in the initiation complex FimD–FimC–FimH, or FimG in the elongation complex FimD–FimC–FimF–FimG–FimH.

¹Institute of Structural and Molecular Biology, University College London and Birkbeck College, Malet Street, London WC1E 7HX, UK. ²The Howard Hughes Medical Institute and Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ³Center for Women's Infectious Disease Research and Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63011, USA.

*These authors contributed equally to this work.

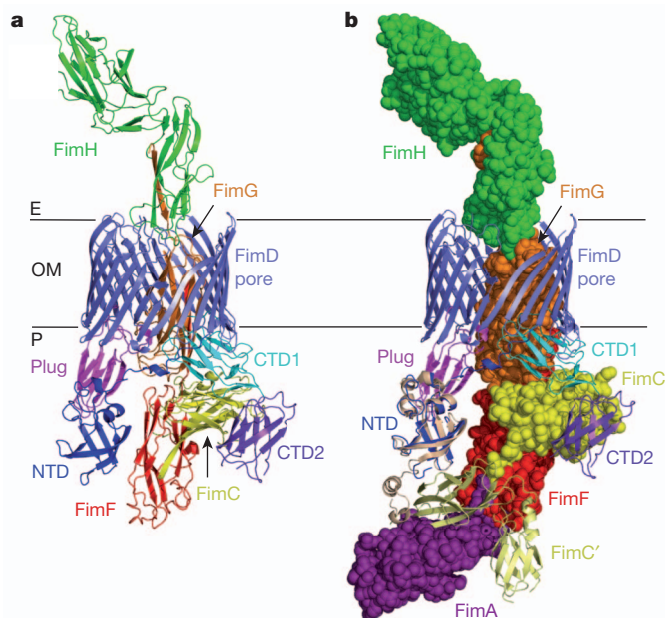


Figure 1 | Structure of FimD-FimC-FimF-FimG-FimH. **a**, Crystal structure of FimD-FimC-FimF-FimG-FimH. All proteins are in ribbon representation: FimD NTD, blue; β -barrel, slate; plug domain, magenta; CTD1, cyan; CTD2, purple; FimC, yellow; FimF, red; FimG, orange; FimH, green. **E**, extracellular; **OM**, outer membrane; **P**, periplasmic. **b**, Recruitment of the next chaperone-subunit complex FimC-FimA in assembly by FimD-FimC-FimF-FimG-FimH. FimD is in ribbon representation, coloured as in **a**. Chaperone FimC (yellow), pilus subunits FimF (red) and FimG (orange), and adhesin FimH (green) are in sphere representation. The recruited chaperone-subunit complex FimC'-FimA (PDB ID, 4DWH; FimC', pale yellow ribbon; FimA, purple spheres) is modelled at the NTD on the basis of the crystal structure of the isolated FimD NTD domain bound to FimC-FimF (PDB ID, 3BWU; see Supplementary Methods for details).

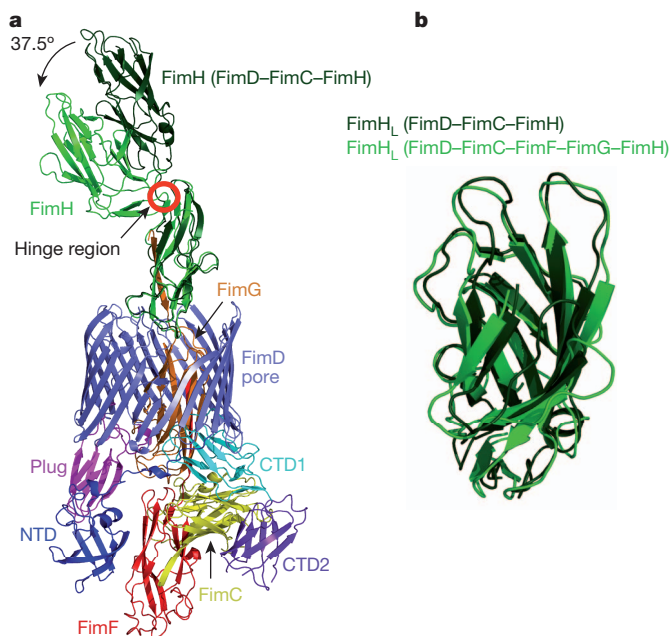


Figure 2 | Comparison of the structures of FimH before and after translocation. **a**, Superposition of the structure of FimH from the initiation complex FimD-FimC-FimH (in dark green) with that of FimH from the elongation complex FimD-FimC-FimF-FimG-FimH (same colour coding as in Fig. 1a). FimH_P was used for the superposition. **b**, Superposition of the structure of FimH_L from the initiation complex (dark green) with that of the same domain from the elongation complex (light green).

In these calculations, the protein assemblies were represented as essentially rigid components. Subunits were moved along a central pore axis (Supplementary Fig. 6a-c) with only minor deviations that can be sampled by rigid-body minimization. This is a reasonable assumption owing to the close steric fit of subunits within the pore that would otherwise clash.

Translating FimG and FimH_L subunits within the pore laterally along a fine grid and plotting the calculated energies reveals a steep energy well, with forces on all sides returning the perturbed subunits to their central, ground-state positions (Methods, Fig. 3a, b and Supplementary Fig. 6d). However, when translated parallel to the pore axis to mimic subunit entry or exit, the calculated energies with FimG in the barrel increase slowly (even for perturbations with a root mean squared deviation of 6 Å), whereas with FimH_L inside the pore there was a steep energy increase (Fig. 3c and Supplementary Fig. 6e). Thus, the subunit or domain inside the pore is more tightly held in the initiation complex than it is in the elongation complex.

In apo-FimD, the plug domain is found inside the pore^{12,13} and must be expelled and replaced by FimH_L in an activation process that might be triggered by engagement of FimH with FimD. When the plug domain (beginning with the structure of apo-FimD (Protein Data Bank (PDB) ID, 3OHN)) was translated towards the periplasm to mimic its extrusion from the pore, the energy increased slowly as it did for FimG (Fig. 3c). Hence, whereas the plug domain and FimG may readily exit the usher pore along the pore axis, FimH_L forms tight interactions. We propose that this may be essential for activation, in which FimH_L displaces the plug domain from inside the pore.

By sweeping a 60°-sector window emanating from the pore axis and computing the binding energy within this

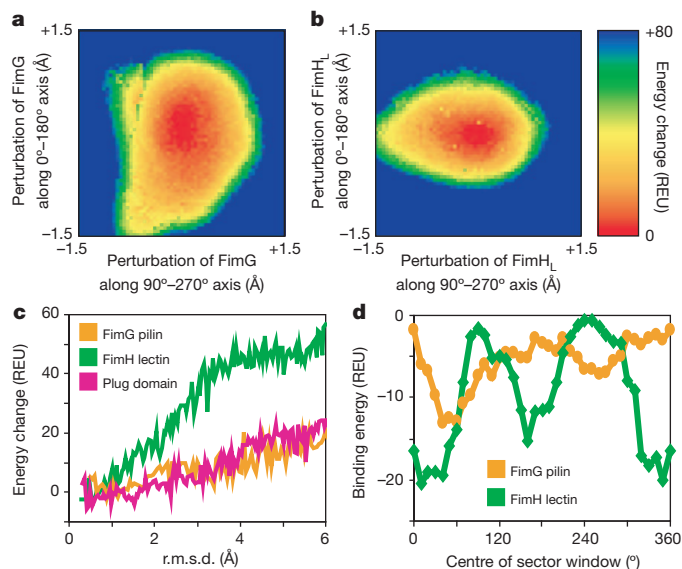


Figure 3 | Steep energy funnels and opposing binding surfaces position the translocating substrate at the centre of the pore. **a**, The heat map corresponding to the view down the pore axis, with the energy increase plotted as FimG is laterally translated (that is, perpendicular to the pore axis) along a finely spaced grid within the usher lumen. FimG located at any position on the grid will experience a force dependent on the slope of the potential energy well, returning FimG to its central ground state. REU, ROSETTA energy units. **b**, As in **a**, with FimH_L now inside the usher lumen. **c**, Subunits or domains occupying the pore were randomly rotated by up to 8° about their geometric centres and translated parallel to the pore axis by up to 6 Å (for the plug domain, displacement was only in the direction of the periplasm to mimic plug extrusion during activation). Calculated energies are plotted against the root mean squared deviation (r.m.s.d.) from the respective minimized crystal structures. Eight thousand perturbations were made for each subunit, with the outline of the lowest-energy conformations shown. **d**, Plot of the respective binding energies as a 60°-sector window emanating from the pore axis is rotated around the FimG pilin structure or the FimH lectin structure.

window (Fig. 3d and Supplementary Figs 7 and 9), we observe two binding-energy troughs and two peaks that are roughly 180° apart for both FimH_L inside the FimD pore of the FimD–FimC–FimH structure and FimG inside the FimD pore of the FimD–FimC–FimF–FimG–FimH structure. Regions of high binding energy correlate with closer atomic contacts as measured by solvent accessibility (Supplementary Fig. 7b), and regions that tightly bind FimG are distinct from those that tightly bind FimH_L, indicating specialized binding sites. Usher pores seem to have evolved binding sites that face each other on opposite sides of the pore, resulting in the placement of the translocating substrate at the very centre of the pore. Because pilin subunits share the same immunoglobulin-like fold and have high sequence similarity (compared with FimG, the pilin domains of FimH, FimF and FimA share 19–25% sequence identity and 62–67% similarity), it is likely that all pilin subunits occupy the same central position in the usher pore, and our results for FimG are generalizable to the rest of these proteins.

It has been previously proposed that incoming chaperone–pilin subunits are recruited to the usher NTD and then transfer after DSE to the usher CTDs during assembly (see above). This requires a rotation of the NTD-bound chaperone–subunit complex by about 100° – 120° after DSE has occurred (Supplementary Fig. 8a). The torque produced by the transfer of the incoming subunit from the NTD to the CTDs might be accommodated through rotations around linker residues between pilus building blocks. Alternatively, rigid-body rotation of the entire tip assembly might necessitate a rotation of the translocating subunit within the pore.

To investigate such a possibility, we calculated an energy landscape as the FimG pilin was rotated in 2° steps and translated in steps of 1 \AA up or down the usher pore. The calculated landscape reveals a continuous low-energy path for FimG as it enters and exits the FimD pore (visualized by following the ‘low-energy’ colours in Fig. 4a and Supplementary Fig. 8) that requires an anticlockwise twist. To predict a possible trajectory through the energy landscape, we introduced a torsional spring term to limit the extent to which FimG can rotate at each translation step (this ensures connectivity in the low-energy path); the entry–exit trajectory for FimG pilin is then derived from the lowest-energy states (Fig. 4b). By repeating these calculations using input structures minimized with and without constraints from the crystallographic electron density, different spring constants and structures refined to a resolution of either 3.8 or 4.1 \AA , a twist of 1.6° – 3.0°

per ångström translation was observed in the trajectories (Supplementary Fig. 8). The presence of a low-energy exit path requiring pilin rotation was further validated with an alternatively defined pore axis based on the membrane-like packing of FimD within the crystal lattice (Supplementary Fig. 9). Visual inspection of the FimD–FimC–FimF–FimG–FimH crystal structure revealed that the three assembled pilin subunits respectively representing entrance into (FimF), residence within (FimG) and exit from (FimH_p) the usher pore are each separated by a translation by 53 \AA and an anticlockwise rotation by 110° or 120° (Supplementary Fig. 8a). Hence, the twist undergone by FimG pilin as it enters or exits the pore and that is captured in the computed energy landscape matches the magnitude and direction of the rotations observed and anticipated from static crystallographic evidence. These results are probably generalizable to other pilin subunits owing to high sequence and structural similarity.

Similar analysis was applied to FimH_L within the FimD lumen, using the FimD–FimC–FimH crystal structure, but no single, clear low-energy entry–exit path was observed (Supplementary Fig. 10). This is not surprising. Indeed, FimH_L differs markedly from all other pilus subunits or domains: whereas FimH_p, FimG, FimF and FimA have a classical pilin fold that consists of six β -strands complemented by the donor strand of the next subunit in assembly, FimH_L has a β -barrel ‘jelly roll’ fold. Furthermore, as shown here, the FimD pore forms tighter interactions with FimH_L than with a representative pilin subunit. Thus, whereas all other pilus subunits and domains are easily extracted from and guided through the pore, unbinding and translocation of FimH_L will require more energy, probably through conformational changes yet to be characterized.

The crystal structure presented here provides unprecedented structural evidence supporting the model first proposed in ref. 3 for the usher-mediated subunit-incorporation cycle during pilus biogenesis. It also demonstrates that usher pores form finely tuned, circular protein–protein interfaces specifically geared to maximize transport, by forming diametrically opposed binding sites that position substrates at the very centre of the pore, and to facilitate specific steps during transport, by imposing rotational and translational constraints through defined energy paths. We anticipate that all transporters will demonstrate similar levels of sophistication by having evolved function-specific features facilitating defined steps during substrate translocation.

METHODS SUMMARY

Expression and purification of FimD–FimC–FimF–FimG–FimH. *Escherichia coli* Tuner cells (Novagen) were transformed with plasmids pNH237 encoding *fimC_{HIS}FGH* under arabinose control and pAN2 encoding *fimD_{Strep}* under isopropylthiogalactoside control. The complex FimD–FimC–FimF–FimG–FimH was then expressed and purified as described in ref. 3.

Crystallization of FimD–FimC–FimF–FimG–FimH. Crystals of the complex were grown in hanging drops at 20°C using the vapour diffusion method, where $1 \mu\text{l}$ of the complex (8 mg ml^{-1}) was mixed with an equal volume of 1.2 – 2.0 M sodium formate.

Structure determination of FimD–FimC–FimF–FimG–FimH. Data were collected at the Diamond Light Source (beamline IO2) at 100 K and were indexed, integrated and scaled to a resolution of 3.8 \AA using the XDS software package¹⁴. A high-resolution cut-off for the data was chosen following ref. 15. The space group was determined using POINTLESS¹⁶. The space group, cell dimensions and data collection statistics are reported in Supplementary Table 1. The phase problem was solved by molecular replacement using PHASER and the structures of FimD (PDB ID, 3RFZ), FimC–FimF (3BWU) and FimG–FimH (3JWN) as search models¹⁷. The molecular replacement model for FimD–FimC–FimF–FimG–FimH obtained by PHASER was refined by alternating rounds of model building with COOT¹⁸ using B-factor-sharpened maps generated with PHENIX¹⁹ and refinement cycles using CCP4 REFMAC²⁰, CNS²¹ and PHENIX¹⁹. An initial rigid-body refinement of the model was carried out using REFMAC. For the next round of refinement, we followed the CNS protocol for low-resolution refinement, which includes torsion angle annealing, grouped B factor and deformable elastic network refinement^{21,22}. Deformable elastic network parameters γ and ω were set to 0.2 and 100 , respectively²³. Final refinement was carried out with PHENIX. The atomic displacement parameters and stereochemistry weights were optimized during the refinement and resulted in tight restraints. Finally, R_{free} converged to 29.8% (Supplementary

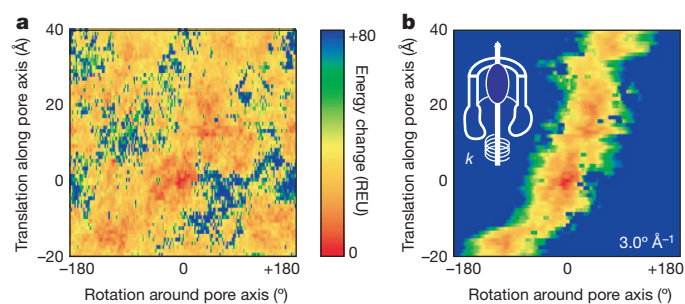


Figure 4 | A low-energy pathway through the pore lumen facilitates translocation of subunits and their transfer from NTD to CTDs. **a**, Starting from the native crystal structure with FimG and its complementing strand from FimF inside the FimD pore (subunits FimH, FimF and FimC were not considered), FimG was translated, in steps of 1 \AA , up to 40 \AA out of the pore and up to 20 \AA back towards the periplasm, along the pore axis. At each translational step, FimG was rotated around the pore axis in 2° steps. Each sampled FimG rotation–translation conformation was minimized after rotamer repacking to resolve small clashes, and the calculated energy for each perturbation is plotted. Lower, more favourable, energies are hotter colours. **b**, The energy landscape calculated in **a** after modification by addition of a torsional spring potential ($E = k(\Delta\theta)^2/2$; represented by the coil with spring constant k twisted around the pore axis in the sketch at top left) to derive a connected trajectory, with $\Delta\theta$ the angle from the lowest-energy FimG conformation before each step of 1 \AA , and starting with FimG positioned at 0 \AA , 0° .

Tables 1 and 2 and Supplementary Figs 3 and 4). The same R_{free} flag was maintained for cross-validation throughout the refinement process.

Computational methods. Before binding-energy calculations were performed, crystal structure coordinates were relaxed in the ROSETTA force field using FASTRELAX with electron density constraints^{11,24} (Methods and Supplementary Figs 11 and 12). FimD sectors were generated on the basis of the angle a residue's C α atom makes with FimD–D208 (set to 0°) when projected on a plane perpendicular to the pore axis. Only FimD and the subunit within the pore were considered (FimG pilin (amino acids 12–144)/FimF Nte (1–12) or FimH_L (1–157)), and calculations of binding energy and buried solvent-accessible surface area were made using ROSETTASCRIPTS²⁵ (Supplementary Fig. 13). Translational and rotational perturbations of the translocating subunit were made with the MMTSB suite²⁶ followed by rotamer repacking and side-chain and backbone gradient minimization in ROSETTA (Supplementary Figs 14–16).

Full Methods and any associated references are available in the online version of the paper.

Received 2 October 2012; accepted 14 February 2013.

- Waksman, G. & Hultgren, S. J. Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nature Rev. Microbiol.* **7**, 765–774 (2009).
- Nishiyama, M., Ishikawa, T., Rechsteiner, H. & Glockshuber, R. Reconstitution of pilus assembly reveals a bacterial outer membrane catalyst. *Science* **320**, 376–379 (2008).
- Phan, G. *et al.* Crystal structure of the FimD usher bound to its cognate FimC–FimH substrate. *Nature* **474**, 49–53 (2011).
- Sauer, F. G. *et al.* Structural basis of chaperone function and pilus biogenesis. *Science* **285**, 1058–1061 (1999).
- Sauer, F. G., Pinkner, J. S., Waksman, G. & Hultgren, S. J. Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **111**, 543–551 (2002).
- Choudhury, D. *et al.* X-ray structure of the FimC–FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* **285**, 1061–1066 (1999).
- Zavialov, A. V. *et al.* Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* **113**, 587–596 (2003).
- Vetsch, M. *et al.* Pilus chaperones represent a new type of protein-folding catalyst. *Nature* **431**, 329–333 (2004).
- Le Trong, I. *et al.* Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like β -sheet twisting. *Cell* **141**, 645–655 (2010).
- Das, R. & Baker, D. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382 (2008).
- Leaver-Fay, A. *et al.* Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
- Huang, Y., Smith, B. S., Chen, L. X., Baxter, R. H. & Deisenhofer, J. Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC. *Proc. Natl Acad. Sci. USA* **106**, 7403–7407 (2009).
- Remaut, H. *et al.* Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell* **133**, 640–652 (2008).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
- Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
- McCoy, A. J. *et al.* PHASER crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of COOT. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
- Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
- Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
- O'Donovan, D. J. *et al.* A grid-enabled web service for low-resolution crystal structure refinement. *Acta Crystallogr. D* **68**, 261–267 (2012).
- DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
- Fleishman, S. J. *et al.* RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS ONE* **6**, e20161 (2011).
- Feig, M., Karanicas, J. & Brooks, C. L. III. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **22**, 377–395 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by Medical Research Council grant 85602 to G.W. D.B. and E.P. are supported by grant P41 GM103533 from the National Institute of General Medical Studies at the US National Institutes of Health (NIH). S.J.H. was supported by grant AI029549 from the National Institute of Allergy and Infectious Disease at the NIH. We thank the staff of beamline ID23-1 at the European Synchrotron Radiation Facility, the staff of beamline IO2 at the Diamond Light source and A. Cole for technical assistance during data collection.

Author Contributions S.G. and E.P. carried out the crystallographic and computational work, respectively. D.B. and G.W. supervised the work. S.G., E.P., D.B. and G.W. analysed the data. S.G., E.P., S.J.H., D.B. and G.W. wrote the paper.

Author Information The atomic coordinates and structure factors of FimD–FimC–FimF–FimG–FimH have been deposited in the Protein Data Bank under accession ID 4J30. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.W. (g.waksman@ucl.ac.uk) or D.B. (dabaker@u.washington.edu).

METHODS

Expression and purification of FimD–FimC–FimF–FimG–FimH. *Escherichia coli* Tuner (Novagen) was transformed with plasmids pNH237 encoding *fimC_{His}FGH* under arabinose control and pAN2 encoding *fimD_{Strep}* under isopropylthiogalactoside (IPTG) control^{13,13}. Bacteria were grown in TB media containing 35 mg ml^{−1} kanamycin and 25 mg ml^{−1} chloramphenicol at 37 °C. Protein overexpression was induced by addition of 100 μM IPTG and 0.1% (w/v) L-arabinose supplemented with 0.1% (v/v) glycerol at $D_{600\text{ nm}}$ 1.0 for 48 h at 16 °C. By not providing the FimA subunit to the assembly system, pilus production stalls after incorporation of the last tip subunit, FimF, resulting in production of homogeneous FimD–tip complexes.

Protein purification of FimD–FimC–FimF–FimG–FimH. FimD–FimC–FimF–FimG–FimH was purified as described in ref. 3 for FimD–FimC–FimH.

Crystallization of FimD–FimC–FimF–FimG–FimH. Plate-like crystals of FimD–FimC–FimF–FimG–FimH were grown in hanging drops at 20 °C using the vapour diffusion method, where 1 μl of the complex (8 mg ml^{−1}) was mixed with an equal volume of 1.2–2.0 M sodium formate and equilibrated against 600 μl reservoir solution. Crystals were transferred into mother liquor with 25% (v/v) glycerol as cryoprotectant and then flash-frozen in liquid nitrogen.

Structure determination of FimD–FimC–FimF–FimG–FimH. Data were collected at the Diamond Light Source (beamline IO2) at 100 K and were indexed, integrated and scaled to a resolution of 3.8 Å using the XDS software package¹⁴. A high-resolution cut-off for the data was chosen following ref. 15 (Supplementary Table 1). The space group was determined using POINTLESS¹⁶. The space group, cell dimensions and data collection statistics are reported in Supplementary Table 1.

The phase problem was solved by molecular replacement using PHASER and the structures of FimD (PDB ID, 3RFZ), FimC–FimF (3BWU) and FimG–FimH (3JWN) as search models¹⁷. The molecular replacement model for FimD–FimC–FimF–FimG–FimH obtained by PHASER was refined by alternating rounds of model building with COOT¹⁸ using B-factor-sharpened maps generated with PHENIX¹⁹ and refinement cycles using CCP4 REFMAC²⁰, CNS²¹ and PHENIX¹⁹. Initial rigid-body refinement (rigid bodies were assigned for subunits FimC, FimF, FimG and FimH and for the β-barrel, the plug and the N-terminal and two C-terminal domains of FimD) was carried out using REFMAC and resulted in an R_{free} value of 45%. For the next round of refinement, we followed the CNS protocol for low-resolution refinement, which includes torsion angle annealing, grouped B factor and deformable elastic network (DEN) refinement^{21,22}. Two B factors were assigned to each amino acid, one to the main-chain atoms and one to the side-chain atoms. FimC (PDB ID, 3BWU), FimD (3RFZ), FimF (3BWU), FimG (3JWN) and FimH (3JWN) were used as DEN reference files. After a two-dimensional grid search for the DEN parameters using the SBGrid Science Portal web service, DEN parameters γ and ω were set to 0.2 and 100, respectively²³. DEN parameter γ balances the influences of the diffraction data and the reference model, whereas ω is the weighting factor for the DEN potentials. The starting annealing temperature was also optimized using the SBGrid Science Portal DEN web service, and was set to 1,500 K. In the final rounds of DEN refinement, the starting and reference models were the same. DEN refinement changed the positioning of the N- and C-terminal domains, which seem to be quite flexible, as well as the shape of the usher β-barrel, and therefore improved R_{free} considerably. DEN refinement converged to $R_{\text{free}} = 31\%$. The final refinement rounds were carried out with PHENIX. The refinement protocol included three rounds of Cartesian and individual B-factor refinement. The atomic displacement parameters and stereochemistry weights were optimized during the refinement and resulted in tight restraints. Finally, R_{free} converged to 29.8%. The same R_{free} flag was maintained for cross-validation throughout the refinement process. Note

that high-resolution data to a resolution of 3.8 Å were included because the resulting final model had a better R_{free} value, improved stereochemistry and improved electron density (Supplementary Tables 1 and 2 and Supplementary Figs 3 and 4).

Computational methods. For analysis of static structures, the refined crystal structures were relaxed in the ROSETTA force field using two rounds of FASTRELAX and Cartesian space minimization (F. DiMaio, manuscript in preparation) with electron density constraints²⁴. B-factor-sharpened $2mF_o - DF_c$ density maps were generated in PHENIX¹⁹ using deposited structure factors for FimD–FimC–FimH or structure factors to a resolution of 4.1 Å for FimD–FimC–FimF–FimG–FimH. ROSETTA command line and score term weights are detailed in Supplementary Fig. 11. The electron-density-constraining score term accounts for ~40% and ~55% of the total score for FimD–FimC–FimF–FimG–FimH and FimD–FimC–FimH, respectively. Relaxed decoys from 50 relaxation runs were well converged both in total score (scores ranged from −5,964 to −5,911 for FimD–FimC–FimF–FimG–FimH and from −6,376 to −6,326 for FimD–FimC–FimH) and structure (root mean squared deviation from an arbitrary reference decoy was ≤ 0.22 Å for FimD–FimC–FimF–FimG–FimH and ≤ 0.14 Å for FimD–FimC–FimH). The lowest-scoring decoys (Supplementary Fig. 12 shows fits of relaxed structures to electron density) were chosen for final analysis in Fig. 3 and Supplementary Figs 7, 8g, h and 9e–h. The ROSETTA correlation coefficient for measuring agreement between coordinates and a density map²⁴ decreased by 3% for FimD–FimC–FimF–FimG–FimH and by 2% for FimD–FimC–FimH following relaxation, indicating a small trade-off between minimizing structural features within the ROSETTA force field and maintaining a close fit to the experimental electron density map.

To generate FimD sectors, residues were included if the C α atom's projection on a plane perpendicular to the pore axis fell within the sector's desired angle range, with the C α projection of FimD–D208 set to 0°. This creates a PDB file containing FimG pilin (amino acids 12–144)/FimF Nte (1–12) or FimH lectin (1–157) bound to only a subset of FimD residues. Binding energy and buried solvent-accessible surface area were calculated using ROSETTASCRIPTS (Supplementary Fig. 13) without rotamer repacking in either the bound or unbound state²⁵ (rotamer repacking at the edges of a FimD sector would be meaningless). We confirmed that the combined binding energy from any set of sectors adding up to a full 360° sweep consistently approximates the binding energy calculated for the complete FimD protein bound to its transported substrate.

Translational and rotational perturbations were made using the convpdb.pl application in the MMTSB suite²⁶. Perturbed structures were minimized in ROSETTA using full-atom score12 with rounds of rotamer repacking, and side-chain and backbone gradient minimization^{11,25} (Supplementary Figs 14–16). In these cases, where large dynamic subunit motions were explored (Figs 3c and 4 and Supplementary Figs 6, 8b–f, 9i and 10), starting structures were also first pre-minimized in full-atom score12, ensuring that any measurements of root mean squared deviation from the initial ground state accurately reflected the applied perturbations, not simply differences from minimizing in different score functions.

For determining potential FimG trajectories through the FimD pore, noise in the energy landscape was dampened by averaging the energy at each sampled FimG rotation–translation with the energies calculated for FimG rotations within 4°. Starting with FimG at 0° and 0 Å, a torsional spring potential ($E = k(\Delta\theta)^2/2$) was added to the energy calculated using ROSETTA. The angle $\Delta\theta$ was measured relative to the lowest-energy conformation in the previous translational step.

Structure figures were generated using The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC (<http://www.pymol.org/>).

Structural basis for the drug extrusion mechanism by a MATE multidrug transporter

Yoshiki Tanaka^{1,2†}, Christopher J. Hipolito³, Andrés D. Maturana⁴, Koichi Ito⁵, Teruo Kuroda⁶, Takashi Higuchi³, Takayuki Katoh³, Hideaki E. Kato^{1,2}, Motoyuki Hattori^{1,2,7}, Kaoru Kumazaki^{1,2}, Tomoya Tsukazaki^{1,2,7}, Ryuichiro Ishitani^{1,2}, Hiroaki Suga³ & Osamu Nureki^{1,2}

Multidrug and toxic compound extrusion (MATE) family transporters are conserved in the three primary domains of life (Archaea, Bacteria and Eukarya), and export xenobiotics using an electrochemical gradient of H⁺ or Na⁺ across the membrane^{1,2}. MATE transporters confer multidrug resistance to bacterial pathogens^{3–6} and cancer cells⁷, thus causing critical reductions in the therapeutic efficacies of antibiotics and anti-cancer drugs, respectively. Therefore, the development of MATE inhibitors has long been awaited in the field of clinical medicine^{8,9}. Here we present the crystal structures of the H⁺-driven MATE transporter from *Pyrococcus furiosus* in two distinct apo-form conformations, and in complexes with a derivative of the antibacterial drug norfloxacin and three *in vitro* selected thioether-macrocyclic peptides, at 2.1–3.0 Å resolutions. The structures, combined with functional analyses, show that the protonation of Asp 41 on the amino (N)-terminal lobe induces the bending of TM1, which in turn collapses the N-lobe cavity, thereby extruding the substrate drug to the extracellular space. Moreover, the macrocyclic peptides bind the central cleft in distinct manners, which correlate with their inhibitory activities. The strongest inhibitory peptide that occupies the N-lobe cavity may pave the way towards the development of efficient inhibitors against MATE transporters.

Cellular export of toxic compounds is an essential process to maintain life^{10–12}. MATE transporters function in the efflux of endogenous cationic, lipophilic substances and xenobiotics, and are ubiquitously distributed in Archaea, Bacteria and Eukarya. Bacterial MATE transporters confer multidrug resistance to pathogens, such as multi-antibiotic-resistant *Staphylococcus aureus*, which has attracted broad attention as a nosocomial infection. In cancer cells, MATE transporters export structurally diverse anti-cancer drugs, causing fatal reductions in the therapeutic efficacies of those drugs. Thus, the demand for the discovery of molecules capable of antagonizing the functions of MATE transporters is high because of their clinical importance, but the screening campaigns for such drug leads have achieved very little success.

The MATE transporter NorM was identified from *Vibrio parahaemolyticus* as a Na⁺/drug antiporter^{13–15}. The crystal structures of NorM from *Vibrio cholerae* (NorM-VC) in the outward-open conformation was reported at 3.65 Å resolution¹⁶. This structure consists of 12 transmembrane helices (TMs) forming two pseudo-symmetrical lobes, presenting a monovalent cation-binding site within the cavity in the carboxy (C)-terminal lobe; a rocker-switch mechanism similar to that used by the major facilitator superfamily transporters¹⁷ was proposed. However, owing to the limited resolution, the structural mechanism of the cation influx and the drug efflux coupling, as well as that allowing the recognition of a broad range of substrates, remained elusive.

A fluorescence-based analysis showed that the MATE from *P. furiosus* (PfMATE) is driven by a H⁺ gradient, unlike NorM (Supplementary Discussion). PfMATE was crystallized using the lipidic cubic phase (LCP) method¹⁸. The crystals grown in LCP diffracted X-rays to 2.1 Å resolution at maximum. The co-crystallization with the cyclic peptide MaL6 (generated by the random non-standard peptide integrated discovery (RaPID) system^{19–22}, see below) improved the quality of the SeMet-derivatized crystals, which facilitated the phase determination by the single anomalous diffraction (SAD) method. Finally, we determined the outward-open apo structures in two different conformations ('straight' and 'bent' conformations at 2.4 and 2.5 Å resolutions, respectively) (Fig. 1a).

The structure of PfMATE consists of an N-lobe (TM1–TM6) and a C-lobe (TM7–TM12), which are related by a pseudo-two-fold symmetry axis. The PfMATE structure adopts a V-shaped conformation, with the central cleft open towards the extracellular side, representing an outward-open state. Thus, the overall conformation of H⁺-driven PfMATE is similar to that of Na⁺-driven NorM-VC²³, which shares approximately 22% sequence identity with PfMATE (Supplementary Figs 1 and 2 and Supplementary Discussion). A large, hydrophobic central cleft, formed between the N- and C-lobes, can be divided into two cavities, the N- and C-lobe cavities. The N-lobe cavity is larger than the C-lobe cavity (Fig. 1a). In NorM-VC, Glu 255, Phe 259, Tyr 367, Asp 371 and Phe 429 recognize Rb⁺ or Cs⁺ in the C-lobe cavity (Supplementary Fig. 2b)¹⁶. These residues are not conserved in PfMATE, in which Met 260, Phe 279 and Trp 283 occupy the cation-binding space (Supplementary Fig. 2c). Therefore, different cation-conducting pathways are probably used by the H⁺- and Na⁺-driven MATE transporters.

Although the present structures of PfMATE are essentially in the outward-open state, we found that it adopts two distinct conformations, 'bent' and 'straight', in terms of the structure of the TM1 helix in the N-lobe (Fig. 1b, c and Supplementary Fig. 3a). The bent-form crystals appeared around pH 6.0, whereas those of the straight form were obtained at pH 7.0–8.0. In the straight conformation, TM1 forms a single, straight helix, whereas in the bent conformation it is kinked at Pro 26 and Gly 30, and bent towards the TM2 side (Fig. 1b and Supplementary Fig. 3b, c). Furthermore, in the bent conformation, the extracellular halves of the TM5 and TM6 helices and their connecting loop are slightly shifted outwards by 9.9°, and thus adopt a more outward-open conformation (Fig. 1a and Supplementary Movie). As a consequence, the N-lobe cavity, which is mainly formed between TM1 and TM2, is collapsed in the bent conformation (Fig. 1c and Supplementary Movie). To investigate the importance of Pro 26, we mutated Pro 26 to Ala and Ile, which reduced the EtBr and H⁺

¹RIKEN Advanced Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan. ²Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. ³Department of Chemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Bunkyo-ku, Tokyo 113-0033, Japan. ⁴Department of Bioengineering Sciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan. ⁵Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8562, Japan. ⁶Department of Genome Applied Microbiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Tsushima, Okayama 700-8530, Japan. ⁷Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. [†]Present address: Department of Medical Chemistry and Cell Biology, Faculty of Medicine, Kyoto University, Konoe-cho, Yoshida, Sakyo-ku, Kyoto, 606-8501, Japan.

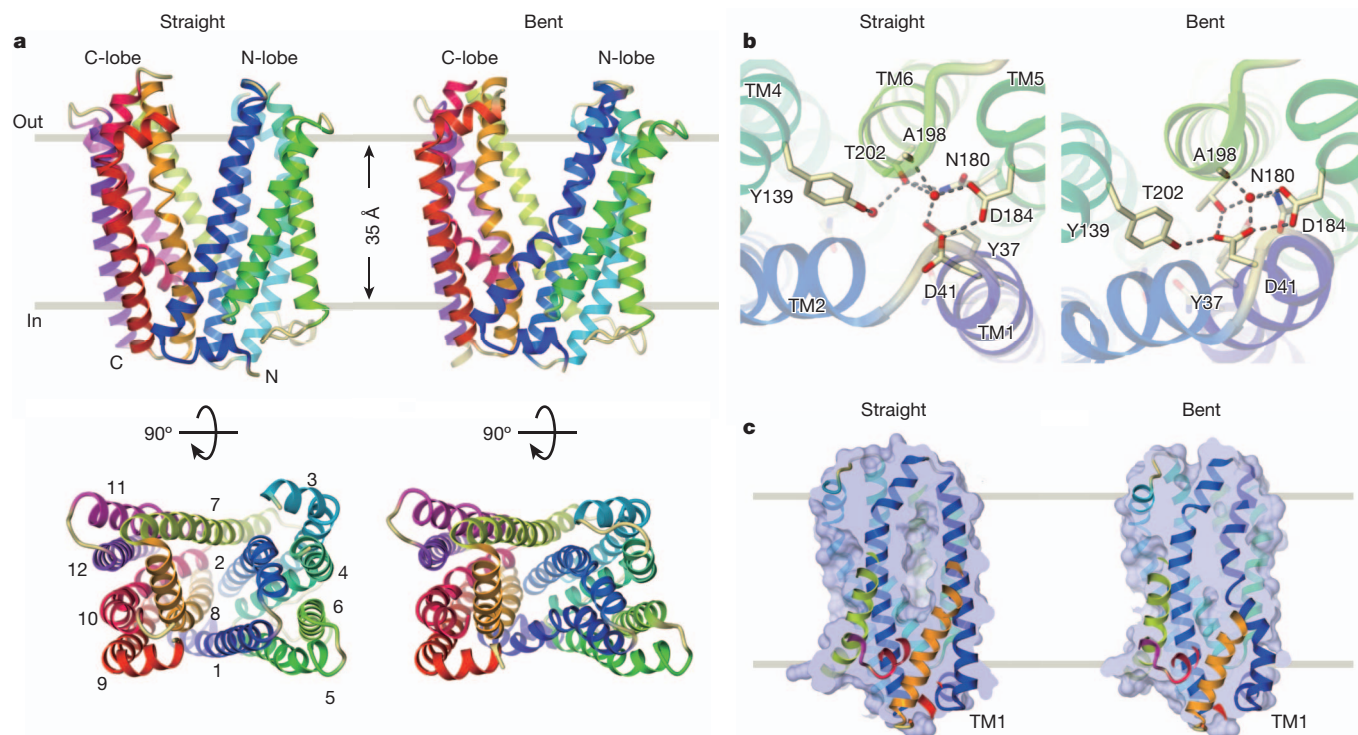


Figure 1 | Overall structures of PfMATE. **a**, Cartoon representations of the two conformations of PfMATE, viewed from the membrane and extracellular sides, with helices TM1–TM12 marked. The molecules are shown as ribbons with each helix coloured as a rainbow, from the N terminus (blue) to the C terminus (purple).

transport activities in the fluorescence-based analysis (Fig. 2a). On the other hand, the P26G and G30A mutants showed normal antiport activities (Fig. 2a). The results of the complementation assay also showed that the P26A mutant could not complement the norfloxacin efflux activity (Fig. 2b). The membrane expression of all of the mutants was confirmed (Supplementary Fig. 4a). We also determined the crystal structure of the P26A mutant at pH 6.5. The structure adopts the straight conformation even under the low-pH conditions, supporting the importance of Pro 26 for the bending of TM1 (Supplementary Fig. 5). Thus, TM1 is kinked at Pro 26 in the bent conformation, which is crucial for the drug export activity.

A drastic rearrangement of the side-chain interactions was also observed between the straight and bent conformations, near the N-lobe cavity. Hydrophilic and acidic residues are clustered at the apex of the N-lobe cavity (Fig. 1b). In the straight conformation, the side chain of Asp 184 (TM5) forms water-mediated hydrogen bonds with Tyr 37, Asn 180 and Thr 202 (Fig. 1b and Supplementary Fig. 3c). Moreover, the O δ atom of Asp 184 is within hydrogen-bonding distance (2.8 Å) with the O δ atom of Asp 41 (TM1), suggesting that the carboxylate group of either Asp 41 or Asp 184 is protonated. Given that Asp 184 protrudes into the protein core, whereas Asp 41 is exposed to the bulk solvent, it is likely that Asp 184 is protonated and Asp 41 is deprotonated. In contrast, in the bent conformation, Asp 41 (TM1) is closer to TM2, and its carboxylate group forms direct hydrogen bonds with Tyr 139 (TM4) and Thr 202 (TM6) (Fig. 1b and Supplementary Fig. 3c). Asp 184 (TM5) again forms a direct hydrogen bond with Asp 41. Thus, a tight water-mediated hydrogen-bond network between the side chains of Asp 41, Asn 180, Asp 184 and Thr 202 and the main-chain carbonyl of Ala 198 (TM6) is formed (Fig. 1b and Supplementary Fig. 3c). This interaction network is further surrounded by the hydrophobic environment, and sequestered from the solvent region of the extracellular half channel (Supplementary Fig. 3d). Therefore, in the bent conformation, the Asp 41 and Asp 184 side chains are probably both protonated, because a charged group in such

b, Close-up views of the interaction networks in the two forms at the apex of the N-lobe cavity. Hydrogen bonds are indicated by black dashed lines. **c**, Surface models viewed from the central cleft, highlighting the N-lobe cavity. All molecular graphics were created with the program CueMol (<http://www.cuemol.org/>).

a low-dielectric environment inside the protein is energetically unfavourable. This is consistent with the fact that the bent-form crystals appeared under acidic conditions. Moreover, a new hydrogen bond is formed between the side chains of Gln 34 (TM1) and Asn 154 (TM4) (Supplementary Fig. 3c). To investigate the importance of these acidic and hydrophilic residues in the N-lobe cavity, we performed fluorescence-based and complementation assays of the Asp 41 and Asp 184 mutants. These mutations abolished both the drug and H⁺ transport activities, and the complementation activities (Fig. 2 and Supplementary Fig. 4c, d), indicating the importance of these acidic residues for the antiport mechanism. Furthermore, the Ala mutations of Tyr 139 and Asn 180 also significantly reduced the transport activity (Fig. 2 and Supplementary Fig. 4c, d). Taken together, the present structures suggested that only the Asp 184 side chain is protonated in the straight conformation, whereas Asp 41 and Asp 184 are both protonated in the bent conformation. The change in the protonation state of the Asp 41 side chain may trigger the reorganization of the interaction network, thereby inducing the structural conversion between the straight and bent conformations (Supplementary Movie).

In the straight conformation, the lipid or monoolein molecules used in the crystallization bind to the N- and C-lobe cavities, and may mimic hydrophobic drug substrates, whereas the N-lobe cavity is collapsed in the bent conformation, and no density peaks were observed (Fig. 3a and Supplementary Fig. 6). To investigate further the drug recognition mechanism, we determined the 2.9 Å resolution crystal structure of PfMATE in the presence of a norfloxacin-derivative compound (Br-NRF)²⁴, in which the 6-fluoro group of norfloxacin is replaced by bromine. The efflux of this compound by PfMATE was confirmed by the fluorescence-based assay (Supplementary Fig. 4c). In this structure, no lipid-like density was observed in the N-lobe cavity. Instead, an electron density that resembled the shape of the Br-NRF molecule was observed (Fig. 3b, c). The Br-NRF molecule is mainly recognized by shape complementarity, involving the side chains of Gln 34, Tyr 37 (TM1), Asn 153 (TM4), Met 173, Ser 177 (TM5), Thr 202, Ser 205, Met 206, Thr 209 and Ile 213 (TM6). The 4-oxo group, the amine

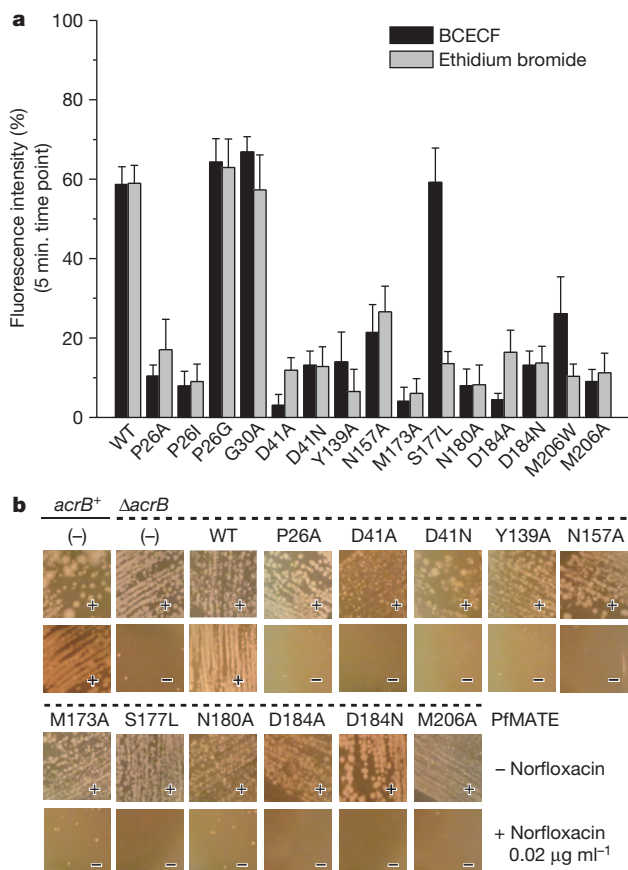


Figure 2 | Mutational analyses of PfMATE. **a**, The proton influx and substrate efflux activities of the PfMATE mutants. The percentages of the fluorescent intensity decrease of BCECF (black) and EtBr (grey) were calculated, as described in the Supplementary Information. The bar graph represents the percentage fluorescence per period of time, namely 5 min after the addition of MES. Error bars, s.d. from experiments repeated five times. WT, wild type. **b**, Growth complementation tests of Δ *acrB* strains harbouring PfMATE mutant vectors, in the presence and absence of quinolone antibacterial agents.

group of the 7-piperazine moiety and the 3-carboxylate group of Br-NRF are further recognized by hydrogen-bond interactions with the side chains of Gln 34 (TM1), Asn 157 (TM4) and Asn 180 (TM5), respectively (Fig. 3c). It should be noted that Tyr 37, Asn 180 and Thr 202 are also involved in the hydrogen-bond network at the apex of the N-lobe cavity, as described above. To explore further the roles of these residues in the substrate recognition, we performed fluorescence-based and complementation assays of their mutants. The Ala mutants of Met 173, Asn 180 and Met 206 abolished both the EtBr and H⁺ transport activities (Fig. 2 and Supplementary Fig. 4c, d), indicating the importance of these residues for the recognition of the drug substrate. Intriguingly, the S177L and M206W mutants retained considerable H⁺ transport activities, but showed significantly reduced EtBr transport activities (Fig. 2), thus showing the decoupling of H⁺ and EtBr transport.

To investigate further the drug efflux mechanism of PfMATE and to identify chemicals that antagonize the PfMATE activity, we performed an *in vitro* selection of thioether-macrocyclic peptides (MaD5, MaD3S and MaL6; Fig. 4a, Supplementary Fig. 7a–c and Supplementary Discussion) that specifically interact with PfMATE, using the RaPID system^{19–22}. We analysed the inhibitory activities of these cyclic peptides by fluorescence-based analyses using *Escherichia coli* cells and outer-membrane-stripped spheroplasts (Fig. 4b and Supplementary Fig. 8d). The PfMATE structures in complex with these cyclic peptides were also determined at 3.0, 2.6 and 2.5 Å resolutions, respectively (Fig. 4c and Supplementary Fig. 7d–f). The MaD5 peptide, which showed the highest inhibitory activity observed in the EtBr-extrusion

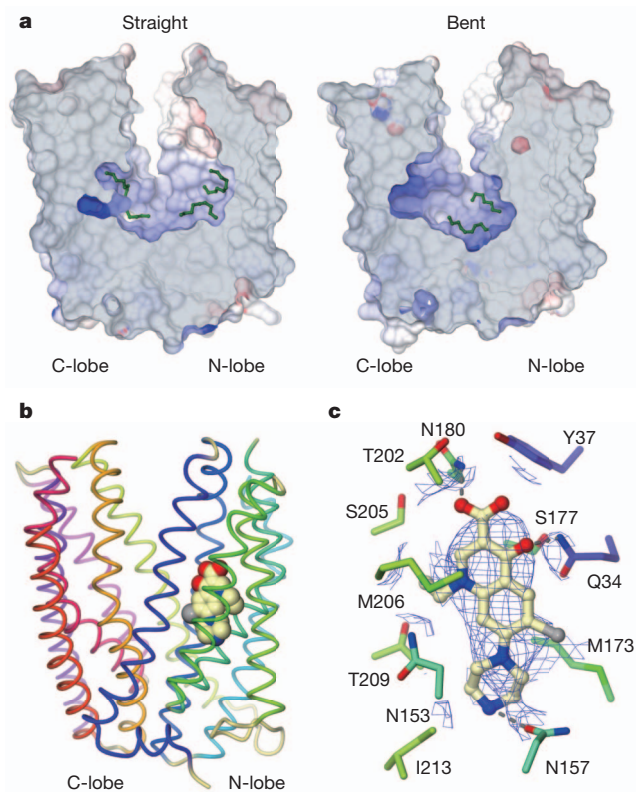


Figure 3 | Complex structure of PfMATE and drug substrate. **a**, Surface model representations of the central clefts in the straight and bent conformations, coloured by the electrostatic potential ranging from blue (+20 kT/e) to red (−20 kT/e). Lipids are shown by green stick models. **b**, The PfMATE structure in the complex with Br-NRF. The bound Br-NRF molecule is shown by a space-filling model. **c**, Drug recognition site, with an unbiased $2mF_o - DF_c$ map contoured at 1.0σ as a mesh within 2.5 \AA . Br-NRF and the surrounding residues are shown by stick models. Hydrogen bonds are indicated by black dashed lines.

fluorescence-monitoring assay (Fig. 4b and Supplementary Fig. 8a), has a lariat-like structure consisting of a seven-amino-acid minicycle head with a 13-amino-acid linear tail (Fig. 4a and Supplementary Fig. 7a, d). In the complex crystal structure, PfMATE adopts the straight conformation, and the MaD5 peptide is deeply bound within the central cleft (Supplementary Fig. 7d). The peptide residues in the minicycle head fill the N-lobe cavity, which is involved in the recognition of the drug substrate. The C-terminal eight residues of the 13-amino-acid-long tail were disordered and not visible in the crystal structure. On the other hand, the MaD3S peptide, which has the same small cyclic head structure but possesses a nine-amino-acid tail with a different sequence composition from the MaD5 tail sequence (Fig. 4a and Supplementary Fig. 7b), showed lower inhibitory activity than that of the MaD5 peptide in the fluorescence-based assay (Fig. 4b and Supplementary Fig. 8b). Therefore, given that the sequences and structures of the minicycle heads of MaD5 and MaD3S are the same, the 13-amino-acid-long tail of MaD5 has an important impact on the inhibitory activity. The long MaD5 tail may restrict the rocker-switch motion of the N- and C-lobes, thereby contributing to the strong inhibitory activity of the MaD5 peptide. Then, to confirm the outer-membrane permeability of the MaD5 peptide, we stained MATE-positive *E. coli* cells using the fluorescent-labelled MaD5 peptide (Supplementary Fig. 9). The result suggested that these anti-MATE inhibitory peptides are able to penetrate the bacterial outer-membrane possibly through porin channels or other passive mechanisms, and reach the MATE protein embedded in the inner-membrane, thereby inhibiting the extrusion function (Fig. 4b and Supplementary Fig. 8). Thus, these peptides may provide new scaffolds for the development of potent inhibitors

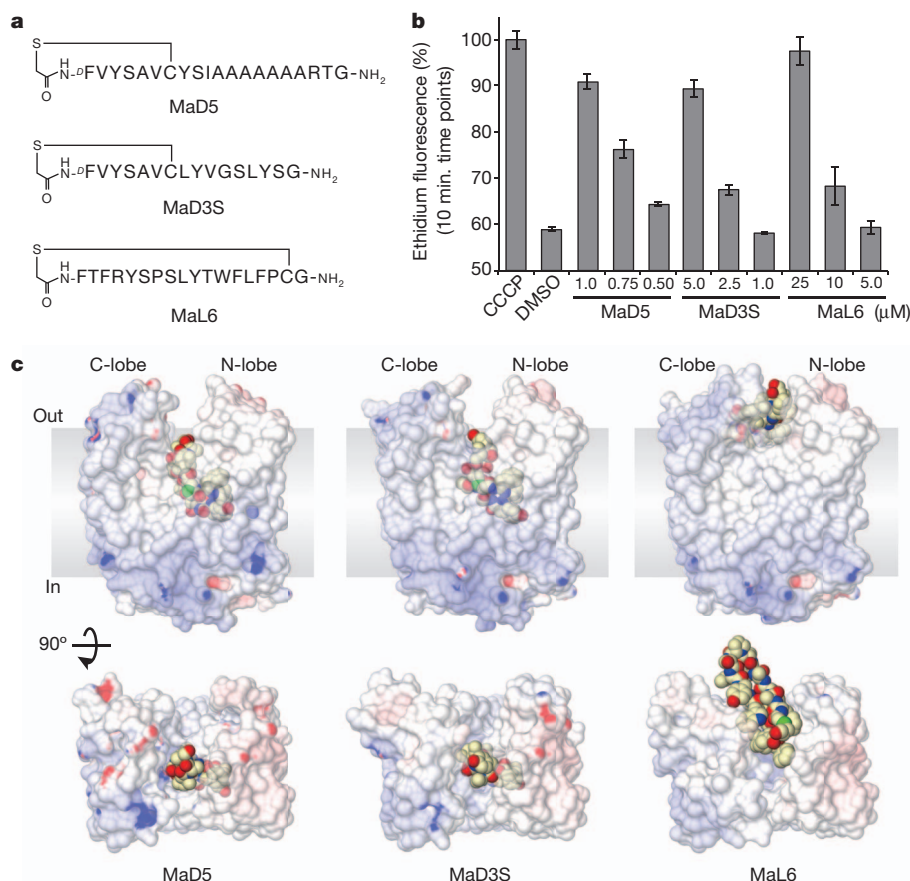


Figure 4 | Complex structures with the macrocyclic peptides. **a**, Schematic representations of the thioether-macrocyclic peptides generated by the RaPID system^{19–22}. ^DF denotes D- α -phenylalanine. **b**, Inhibition of the PfMATE activity by the macrocyclic peptides. *E. coli* C41(DE3) Δ acrB cells expressing PfMATE were treated with different concentrations of peptides as indicated in

the x-axes of graphs. Error bars, s.d. from triplicate experiments. Control is dimethylsulphoxide (DMSO) without peptide. **c**, Surface models of PfMATE complexes with the macrocyclic peptides, viewed from the membrane (upper panels) and extracellular side (lower panels). The bound peptides are depicted by space-filling models.

against MATE transporter families from bacteria and eukaryotes, thus affecting the treatments of multidrug-resistant pathogens and cancer cells.

The present analyses allowed us to propose the last step of the H⁺-driven drug extrusion mechanism by MATE (Supplementary Fig. 10 and Supplementary Movie). Immediately after the structural conversion from the inward-open to outward-open state, the transporter is deprotonated, and only the drug substrate is bound to the transporter (state 1 in Supplementary Fig. 10, corresponding to the present Br-NRF complex structure). After the protonation of Asp 41, its side chain becomes sequestered within the hydrophobic environment, thereby inducing the bending of TM1 at Pro 26, which collapses the N-lobe cavity and extrudes the bound drug substrate into the extracellular space (state 2 in Supplementary Fig. 10, corresponding to the present bent-form structure). Finally, the protonated and substrate-unloaded transporter undergoes further structural conversion to the inward-open state (states 3–5 in Supplementary Fig. 10). During the review process, the crystal structures of Na⁺- and substrate-bound NorM from *Neisseria gonorrhoeae* were reported at 3.5–3.6 Å resolutions²⁵, which may represent the intermediate state during the extrusion of the substrate on the basis of our above mechanism. The complete understanding of the H⁺/drug antiport cycle of MATE awaits the elucidation of the structures capturing the occluded and inward-open states.

METHODS SUMMARY

PfMATE was expressed in *E. coli* C41 cells. The cells were disrupted, and the membrane fraction was collected by ultracentrifugation. The protein was solubilized with DDM and purified by Ni²⁺-affinity chromatography, followed by gel-filtration chromatography in the presence of Cymal-6. PfMATE was mixed with

liquefied monoolein and crystallized in reservoir solutions (28–32% PEG400, 50 mM MES-NaOH, 20 mM CaCl₂, 100 mM NaSCN; pH 6.5). SeMet derivatives were prepared in *E. coli* C41 Met[−] cells. The SeMet protein was mixed with the cyclic-peptide MaL6 and crystallized. X-ray diffraction data sets were collected by the helical data collection method²⁶ on beamline BL32XU at SPring-8 with the approval of RIKEN. The crystal structure of PfMATE was determined by the SAD method.

For the fluorescent analysis, spheroplasts of *E. coli* Δ acrB expressing PfMATE were incubated with 5 μ M 2',7'-bis(2-carboxyethyl)-5(6)-carboxyfluorescein (BCECF)-AM (the acetoxymethyl ester permeates cells, whereas the hydrolysed BCECF cannot permeate the cell membrane and accumulates within the cell)²⁷ and 100 μ M norfloxacin for 30 min at room temperature. The spheroplasts were centrifuged briefly, and placed on glass coverslips under a confocal microscope (FV1000, Olympus). BCECF fluorescence was measured with excitation at 490 nm and emission at 530 nm. To reduce the external pH to 5.5, 10 mM MES was added. The protonophore carbonyl cyanide *m*-chlorophenylhydrazone was then added at 50 μ M, to equilibrate the intracellular and extracellular pH.

The selection of PfMATE-binding cyclic peptides and the chemical synthesis of the peptides were performed according to the RaPID system protocols^{19–22}. The inhibitory activities of the macrocyclic peptides were investigated by EtBr accumulation, according to the literature²⁸. EtBr (50 μ M final concentration) was added to *E. coli* Δ acrB cells expressing PfMATE, in the presence of peptide, and the fluorescence emissions of the controls and the peptide-PfMATE mixtures were immediately measured by a Flexstation 3 (Molecular Devices).

Full Methods and any associated references are available in the online version of the paper.

Received 22 July 2012; accepted 14 February 2013.

Published online 27 March 2013.

1. Brown, M. H., Paulsen, I. T. & Skurray, R. A. The multidrug efflux protein NorM is a prototype of a new family of transporters. *Mol. Microbiol.* **31**, 394–395 (1999).

2. He, G. X. *et al.* An H⁺-coupled multidrug efflux pump, PmpM, a member of the MATE family of transporters, from *Pseudomonas aeruginosa*. *J. Bacteriol.* **186**, 262–265 (2004).
3. Kaatz, G. W., McAleese, F. & Seo, S. M. Multidrug resistance in *Staphylococcus aureus* due to overexpression of a novel multidrug and toxin extrusion (MATE) transport protein. *Antimicrob. Agents Chemother.* **49**, 1857–1864 (2005).
4. McAleese, F. *et al.* A novel MATE family efflux pump contributes to the reduced susceptibility of laboratory-derived *Staphylococcus aureus* mutants to tigecycline. *Antimicrob. Agents Chemother.* **49**, 1865–1871 (2005).
5. Tsuda, M. *et al.* Oppositely directed H⁺ gradient functions as a driving force of rat H⁺/organic cation antiporter MATE1. *Am. J. Physiol.* **292**, F593–F598 (2007).
6. Becker, M. L. *et al.* Genetic variation in the multidrug and toxin extrusion 1 transporter protein influences the glucose-lowering effect of metformin in patients with diabetes: a preliminary study. *Diabetes* **58**, 745–749 (2009).
7. Minematsu, T. & Giacomini, K. M. Interactions of tyrosine kinase inhibitors with organic cation transporters and multidrug and toxic compound extrusion proteins. *Mol. Cancer Ther.* **10**, 531–539 (2011).
8. Ito, S. *et al.* Potent and specific inhibition of mMate1-mediated efflux of type I organic cations in the liver and kidney by pyrimethamine. *J. Pharmacol. Exp. Ther.* **333**, 341–350 (2010).
9. Kusuhara, H. *et al.* Effects of a MATE protein inhibitor, pyrimethamine, on the renal elimination of metformin at oral microdose and at therapeutic dose in healthy subjects. *Clin. Pharmacol. Ther.* **89**, 837–844 (2011).
10. Dawson, R. J. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180–185 (2006).
11. Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* **443**, 173–179 (2006).
12. Wasaznik, A., Grinholc, M. & Bielawski, K. P. Active efflux as the multidrug resistance mechanism. *Postepy Hig. Med. Dosw.* **63**, 123–133 (2009).
13. Morita, Y. *et al.* NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrob. Agents Chemother.* **42**, 1778–1782 (1998).
14. Morita, Y., Kataoka, A., Shiota, S., Mizushima, T. & Tsuchiya, T. NorM of *Vibrio parahaemolyticus* is an Na⁺-driven multidrug efflux pump. *J. Bacteriol.* **182**, 6694–6697 (2000).
15. Otsuka, M. *et al.* Identification of essential amino acid residues of the NorM Na⁺/multidrug antiporter in *Vibrio parahaemolyticus*. *J. Bacteriol.* **187**, 1552–1558 (2005).
16. He, X. *et al.* Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature* **467**, 991–994 (2010).
17. Law, C. J., Maloney, P. C. & Wang, D. N. Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **62**, 289–305 (2008).
18. Cherezov, V. Lipidic cubic phase technologies for membrane protein structural studies. *Curr. Opin. Struct. Biol.* **21**, 559–566 (2011).
19. Hipolito, C. J. & Suga, H. Ribosomal production and *in vitro* selection of natural product-like peptidomimetics: the FIT and RaPID systems. *Curr. Opin. Chem. Biol.* **16**, 196–203 (2012).
20. Hayashi, Y., Morimoto, J. & Suga, H. *In vitro* selection of anti-Akt2 thioether-macrocyclic peptides leading to isoform-selective inhibitors. *ACS Chem. Biol.* **7**, 607–613 (2012).
21. Yamagishi, Y. *et al.* Natural product-like macrocyclic *N*-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed *de novo* library. *Chem. Biol.* **18**, 1562–1570 (2011).
22. Morimoto, J., Hayashi, Y. & Suga, H. Discovery of macrocyclic peptides armed with a mechanism-based warhead: isoform-selective inhibition of human deacetylase SIRT2. *Angew. Chem. Int. Ed.* **51**, 3423–3427 (2012).
23. Altschul, S. F. *et al.* Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **272**, 5101–5109 (2005).
24. Koga, H., Itoh, A., Murayama, S., Suzue, S. & Irikura, T. Structure-activity relationships of antibacterial 6,7- and 7,8-disubstituted 1-alkyl-1,4-dihydro-4-oxoquinoline-3-carboxylic acids. *J. Med. Chem.* **23**, 1358–1363 (1980).
25. Lu, M. *et al.* Structures of a Na⁺-coupled, substrate-bound MATE multidrug transporter. *Proc. Natl Acad. Sci. USA* **110**, 2099–2104 (2013).
26. Flot, D. *et al.* The ID23–2 structural biology microfocus beamline at the ESRF. *J. Synchrotron Radiat.* **17**, 107–118 (2010).
27. Tsujimoto, K., Semadeni, M., Huflejt, M. & Packer, L. Intracellular pH of halobacteria can be determined by the fluorescent dye 2', 7'-bis(carboxyethyl)-5(6)-carboxyfluorescein. *Biochem. Biophys. Res. Commun.* **155**, 123–129 (1988).
28. Li, X. Z., Poole, K. & Nikaido, H. Contributions of MexAB-OprM and an EmrE homolog to intrinsic resistance of *Pseudomonas aeruginosa* to aminoglycosides and dyes. *Antimicrob. Agents Chemother.* **47**, 27–33 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to the beam-line staff at BL32XU of SPring-8 for assistance in data collection (proposals 2011B1062, 2012A1087 and 2012B1161), and the RIKEN BioResource Center (Ibaraki, Japan) for providing the *P. furiosus* genomic DNA. We are also grateful to N. Dohmae and Y. Sugita (RIKEN Advanced Science Institute, Japan) for discussions. This work was supported by the Japan Society for the Promotion of Science (JSPS) through its 'Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST program)' to O.N.; by the Core Research for Evolutional Science and Technology Program 'The Creation of Basic Medical Technologies to Clarify and Control the Mechanisms Underlying Chronic Inflammation' of Japan Science and Technology Agency to O.N.; and by a Grant-in-Aid for Scientific Research (S) (24227004) and a Grant-in-Aid for Young Scientists (A) (22687007) from MEXT to O.N. and R.I., respectively. This work was also supported by a JSPS Grant-in-Aid for the Specially Promoted Research (21000005) and MEXT Platform for Drug Discovery, Informatics, and Structural Life Science to H.S., and a Grant-in-Aid for JSPS post-doctoral fellows to C.J.H. (P11344).

Author Contributions Y.T. expressed and purified PfMATE for crystallization, collected the diffraction data, solved the structures and made the mutants for functional analyses. C.J.H. performed selections, syntheses and inhibition assays of macrocyclic peptides. A.D.M. performed the fluorescence analysis. K.I. performed growth complementation tests. T.K. performed drug susceptibility tests. T.H. synthesized Br-NRF. K.K. and H.E.K. assisted with data collection. M.H. and T.T. contributed to the early stage of the project. Y.T., C.J.H., H.E.K., R.I. and O.N. wrote the manuscript. H.S. and O.N. directed and supervised all of the research.

Author Information The coordinates and structure factors for the *P. furiosus* apo MATE ('straight' and 'bent' conformations), P26A mutant, Br-NRF-bound MATE and peptide-bound MATEs have been deposited in the Protein Data Bank, under the accession numbers 3VVN, 3VVO, 3W4T, 3VVP, 3VVQ, 3VVR and 3VVS, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.N. (nureki@biochem.s.u-tokyo.ac.jp) and H.S. (hsuga@chem.s.u-tokyo.ac.jp).

METHODS

Cloning, expression and purification. The PfMATE gene was cloned from *P. furiosus* (strain ATCC 43587) and spliced into the expression vector pET11a (Novagen), between the NdeI and XhoI restriction sites of the multiple cloning site. The hexa-histidine tag (His₆-tag) was introduced at the C terminus of the PfMATE construct. The pET11a-PfMATE plasmid was transformed into the *E. coli* C41(DE3) Δ acrB strain, constructed in-house. The cells were grown in a 5 l culture at 37 °C to an attenuation at 600 nm (D_{600}) of 0.5, and the gene expression was induced with 0.5 mM isopropyl β -D-thiogalactopyranoside (IPTG) at 20 °C for 12 h. The cells were pelleted by centrifugation at 4,000g, and were disrupted by a Microfluidizer (Microfluidics). After centrifugation (25,000g), the supernatant was ultra-centrifuged (200,000g), and the membrane fraction was collected. The PfMATE was solubilized from the membrane fraction with *n*-dodecyl- β -D-maltopyranoside (DDM), and was purified by the following chromatography steps. The insoluble material was removed by ultracentrifugation (Beckman Type 70 Ti rotor, 150,000g, 30 min), and the supernatant was mixed with Ni-NTA resin (QIAGEN). The His₆-tag of the PfMATE was cleaved by 1 μ g ml⁻¹ trypsin (Invitrogen) at 4 °C overnight, and the protein was re-chromatographed on a Ni-NTA column. The His₆-tag-cleaved PfMATE was further purified by gel filtration (Superdex 200 10/300 GL, GE Healthcare) in 20 mM HEPES-NaOH, pH 7.0, containing 20 mM NaCl and 0.06% Cymal-6.

Crystallization. The purified protein was concentrated to about 8 mg ml⁻¹ with a centrifugal filter device (Millipore 50 kDa MW cutoff), for crystallization. PfMATE was mixed with liquefied monoolein (Sigma) in a 2:3 protein to lipid ratio (w/w), using the twin-syringe mixing method. Aliquots (100 nl) of the protein-LCP mixture were spotted on a 96-well sandwich plate and overlaid with 1 μ l of precipitant solution by the crystallization robot, mosquito LCP (TTP LabTech). Native crystals for data collection were grown at 20 °C in reservoir solutions (28–30% PEG400, 50 mM MES-NaOH, 20 mM CaCl₂, 100 mM NaSCN; pH 6.0–8.0). The crystals grew to full size in 5–7 days. The crystals were flash-cooled, using reservoir solution containing 32% PEG400 as a cryoprotectant, and were stored in liquid nitrogen. The substrate complex of Br-NRF was dissolved to a 100 mM concentration in 10 mM sodium acetate, pH 4.5, and was mixed with the protein solution at a 20:1 protein to drug ratio (v/v). For the thioether-macrocyclic peptide complexes, the macrocyclic peptides were dissolved in DMSO to a 20 mM concentration, and were mixed with the protein solution in a 20:1 protein to peptide ratio (v/v). Crystallization was performed as described above in reservoir solutions (26–28% PEG550MME, 100 mM Tris-HCl, 100 mM Li₂SO₄; pH 8.0). The P26A mutant was purified by gel filtration in 20 mM MES-NaOH, pH 6.0, containing 20 mM NaCl and 0.06% Cymal-6, and crystallization was performed as described above in reservoir solutions (30% PEG400, 50 mM MES, 100 mM NaSCN, 20 mM CaCl₂; pH 6.5).

Selenomethionine-derivatized PfMATE (SeMet PfMATE) was overexpressed in *E. coli* C41(DE3)Met⁻, and was purified in the same manner as the native protein. The SeMet protein was mixed with the macrocyclic peptide MaL6, and incubated for 1 h at 4 °C. Crystals for data collection were grown at 20 °C by the LCP glass sandwich batch method, using 1.0 μ l reservoir solution (30–32% PEG400, 100 mM MES, 100 mM magnesium acetate; pH 6.5). The co-crystallization with MaL6 improved the quality and reproducibility of the crystals, especially in the crystallization of SeMet PfMATE, and thus facilitated the phase determination by the SAD method.

Data collection and structure determination. The crystal structure of PfMATE was determined by SAD phasing, using the SeMet-derivatized PfMATE-MaL6 complex crystal. X-ray diffraction data sets were collected by the helical data collection method²⁶ on beamline BL32XU at SPring-8 with the approval of RIKEN, using a micro beam with a 1 μ m width and a 5 μ m height²⁹. The diffraction data were processed using the HKL2000 program suite (HKL Research). The heavy atom sites were identified with the program SnB³⁰. The experimental phases were calculated using the program AutoSHARP³¹. The homology model, generated from NorM-VC (Protein Data Bank accession number 3MKU) with the program MODELLER³², was fitted to the initial experimental electron density map and manually modified using the program Coot^{33,34}. The model was refined with the programs Phenix³⁵ and autoBUSTER³⁶. The final models were examined by Ramachandran plots, using the program MolProbity³⁷. The crystal structures of the native MATE and the complexes with the Br-NRF, MaL6, MaD5 and MaD3S peptides were determined by molecular replacement, using the program PHASER^{38,39}. The best data set of the native MATE was refined with $R_{\text{work}}/R_{\text{free}} = 21.7\%/24.4\%$ at 2.4 Å resolution (Supplementary Tables 1 and 2). The asymmetric units of the crystals of the native MATE and the peptide complex contain a single MATE molecule. The asymmetric unit of the drug complex contains two molecules. Molecular graphics were created with the program CueMol (<http://www.cuemol.org/>).

In vivo complementation assays. The drug-sensitive strain (BW25113) Δ acrB was transformed with each pSUIQ plasmid⁴⁰, and the transformants were obtained primarily on lysogeny broth (LB) (7 μ g ml⁻¹ chloramphenicol) plates. Each colony was streaked on 0.02 μ g ml⁻¹ norfloxacin plates, and colony growth was monitored.

Generation of spheroplasts. Giant spheroplasts, which were large enough for fluorescence measurement by confocal microscopy, were generated by genetic and pharmacological procedures^{41,42}. To generate the giant spheroplasts, cells were inoculated in 5 ml of LB medium (50 μ g ml⁻¹ ampicillin) in a 14 ml test tube, and were incubated at 37 °C aerobically by rotation at 200 r.p.m., until D_{600} reached 0.3–0.5. At the desired D_{600} , a 5 ml portion of the exponentially growing culture was diluted tenfold into 5 ml of pre-warmed LB medium, supplemented with 60 μ g ml⁻¹ cephalixin (SIGMA) to block cell division. After 3 h of incubation at 37 °C with rotation at 200 r.p.m., the un-septated filaments were collected by centrifuging the entire culture at 3,000g for 3 min. The filament pellet was then re-suspended in 500 μ l of 0.8 M sucrose, by pipetting. A 30 μ l aliquot of 1 M Tris-HCl (pH 8.0), 24 μ l of 0.25 mg ml⁻¹ lysozyme (300 units, SIGMA L6876), 6 μ l of 5 mg ml⁻¹ DNase I (SIGMA D5025) and 6 μ l 125 mM EDTA-NaOH (pH 8.0) were sequentially added and mixed immediately in between additions, by inverting the tube a few times. After an 8 min incubation at room temperature, a 100 μ l aliquot of stop solution (10 mM Tris-HCl, pH 8.0, 0.7 M sucrose, 20 mM MgCl₂) was added to terminate the cell wall digestion. The spheroplasts were frozen at –20 °C.

BCECF fluorescence analysis. The changes in pH were monitored in spheroplasts, using the pH probe BCECF (Molecular Probes, Invitrogen). In summary, the spheroplasts were incubated with 5 μ M BCECF-AM²⁷ and 100 μ M norfloxacin for 30 min at room temperature. The spheroplasts were then centrifuged briefly (1,500g, 5 min), and were placed on glass coverslips under a confocal microscope (FV1000, Olympus). The spheroplast medium was composed of 150 mM KCl, 20 mM CaCl₂, 5 mM HEPES and 300 mM sucrose (pH 7.5, KOH). BCECF fluorescence was observed with excitation at 488 nm and emission at 530 nm. The fluorescence intensity at time 0 was set as the maximum (100%), to avoid the spheroplast-to-spheroplast variability in fluorescent intensity. Spheroplasts on glass coverslips were randomly selected for fluorescence measurement. To reduce the external pH to 5.5, 10 mM MES was added. After the addition of MES, 50 μ M of the protonophore carbonyl cyanide-*m*-chlorophenylhydrazone (CCCP) was added, to equilibrate the intracellular and extracellular pH. Each graph has the mean and s.e.m. of $n = 8$ –15 spheroplasts (measured in three to six experiments with one to five spheroplasts measured for each experiment).

Ethidium bromide fluorescence analysis. To monitor drug export by MATE, spheroplasts were loaded with ethidium bromide (20 μ M) for 40 min, and then washed by brief centrifugation (1,500g for 5 min). The spheroplasts were suspended in 150 mM KCl, 20 mM CaCl₂, 5 mM HEPES and 300 mM sucrose (pH 7.5, KOH). Ethidium bromide fluorescence was observed with excitation at 490 nm and emission at 600 nm, using a confocal microscope (FV1000, Olympus). To generate a proton gradient, 10 mM MES was added, thus reducing the pH to 5.5. Each graph has the mean and s.e.m. of $n = 9$ –19 spheroplasts (measured in three to six experiments with one to eight spheroplasts measured during each experiment).

Selection for binding cyclic peptides and chemical synthesis. The selection of PfMATE-binding cyclic peptides and the chemical synthesis of the selected peptides were performed according to the RAPID system protocols. Libraries were constructed using a random region consisting of 7–15 NNK codons. Translation initiation of the 'L' library used chloroacetyl-L-phenylalanine-tRNA^{init}_{CAU}. Translation initiation of the 'D' library used chloroacetyl-D-phenylalanine-tRNA^{init}_{CAU}. The selections performed at 4 °C and 37 °C were halted at the seventh and sixth rounds, respectively. The details of the construction and characterization of the PfMATE-binding cyclic peptides will be published in a separate report.

Inhibition assay. The inhibitory activities of the MaL6, MaD5 and MaD3S peptides were tested by an ethidium bioaccumulation assay, according to the literature with modifications²⁸. *E. coli* Δ acrB was used to express PfMATE. Cells containing the PfMATE expression vector were grown in LB medium, containing 100 μ g ml⁻¹ ampicillin, overnight at 37 °C, with shaking at 200 r.p.m. A stationary-phase *E. coli* culture was diluted 100-fold in fresh LB medium containing 100 μ g ml⁻¹ ampicillin. Cells were grown at 37 °C to $D_{600} = 0.5$, and expression was induced by the addition of IPTG (final concentration of 0.5 mM). The induced cells were incubated for 3 h at 37 °C, with shaking at 200 r.p.m. The cells collected from 500 μ l of induced culture were washed once with 50 mM Tris buffer (500 μ l, pH 7.0), and re-suspended in 50 mM Tris buffer (pH 7.0) to a final D_{600} of 0.5. To the wells of a 96-well black plate (1/2 area, Perkin Elmer), a 44.5 μ l portion of the above cell suspension was added to 0.5 μ l of peptide, ranging in concentration from 0 to 2.5 mM in DMSO. For a positive control, 0.5 μ l of a 10 mM solution of CCCP dissolved in DMSO was used, instead of the peptide dissolved in DMSO. EtBr

(0.5 mM, 5 μ l) was added to the *E. coli* cells mixed with the peptide, a blank negative control, or a CCCC positive control, and the plate containing the mixtures was immediately placed in the plate reader (Flexstation 3, Molecular Devices). Note that there was about a 15 s delay between the addition of EtBr and the first reading ($t = 0$ s). Fluorescence readings were taken at 30 s intervals, from 0 to 900 s. Each trial contained 2.2×10^8 *E. coli* cells in a 50 μ l final volume.

29. Hirata, K. *et al.* New micro-beam beamline at SPring-8, targeting at protein micro-crystallography. *AIP Conf. Proc.* **1234**, 901–904 (2010).
30. Xu, H., Smith, A. B., Sahinidis, N. V. & Weeks, C. M. SnB version 2.3: triplet sieve phasing for centrosymmetric structures. *J. Appl. Cryst.* **41**, 644–646 (2008).
31. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
32. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **2**, 2.9.1–2.9.31 (2007).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
34. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
35. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
36. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68**, 368–380 (2012).
37. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
38. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
39. Zwart, P. H. *et al.* Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435 (2008).
40. Ito, K., Ebihara, K., Uno, M. & Nakamura, Y. Conserved motifs in prokaryotic and eukaryotic polypeptide release factors: tRNA-protein mimicry hypothesis. *Proc. Natl Acad. Sci. USA* **93**, 5443–5448 (1996).
41. Martinac, B., Buechner, M., Delcour, A. H., Adler, J. & Kung, C. Pressure-sensitive ion channel in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **84**, 2297–2301 (1987).
42. Kuo, M. M., Saimi, Y., Kung, C. & Choe, S. Patch clamp and phenotypic analyses of a prokaryotic cyclic nucleotide-gated K^+ channel using *Escherichia coli* as a host. *J. Biol. Chem.* **282**, 24294–24301 (2007).

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature11994

Corrigendum: Structure of full-length *Drosophila* cryptochrome

Brian D. Zoltowski, Anand T. Vaidya, Deniz Top, Joanne Widom, Michael W. Young & Brian R. Crane

Nature **480**, 396–399 (2011); doi:10.1038/nature10618

We report an improved model of the *Drosophila* cryptochrome structure that corrects errors in the original coordinates (PDB ID 3TVS). We apologize for these errors. The new coordinates have been deposited in the Protein Data Bank as 4GU5. For further information, please see the accompanying Brief Communication Arising¹.

1. Levy, C. *et al.* Updated structure of *Drosophila* cryptochrome. *Nature* **495**, E3–E4 (2013).

ERRATUM

doi:10.1038/nature12067

Erratum: A sub-Mercury-sized exoplanet

Thomas Barclay, Jason F. Rowe, Jack J. Lissauer, Daniel Huber, François Fressin, Steve B. Howell, Stephen T. Bryson, William J. Chaplin, Jean-Michel Désert, Eric D. Lopez, Geoffrey W. Marcy, Fergal Mullally, Darin Ragozzine, Guillermo Torres, Elisabeth R. Adams, Eric Agol, David Barrado, Sarbani Basu, Timothy R. Bedding, Lars A. Buchhave, David Charbonneau, Jessie L. Christiansen, Jørgen Christensen-Dalsgaard, David Ciardi, William D. Cochran, Andrea K. Dupree, Yvonne Elsworth, Mark Everett, Debra A. Fischer, Eric B. Ford, Jonathan J. Fortney, John C. Geary, Michael R. Haas, Rasmus Handberg, Saskia Hekker, Christopher E. Henze, Elliott Horch, Andrew W. Howard, Roger C. Hunter, Howard Isaacson, Jon M. Jenkins, Christoffer Karoff, Steven D. Kawaler, Hans Kjeldsen, Todd C. Klaus, David W. Latham, Jie Li, Jorge Lillo-Box, Mikkel N. Lund, Mia Lundkvist, Travis S. Metcalfe, Andrea Miglio, Robert L. Morris, Elisa V. Quintana, Dennis Stello, Jeffrey C. Smith, Martin Still & Susan E. Thompson

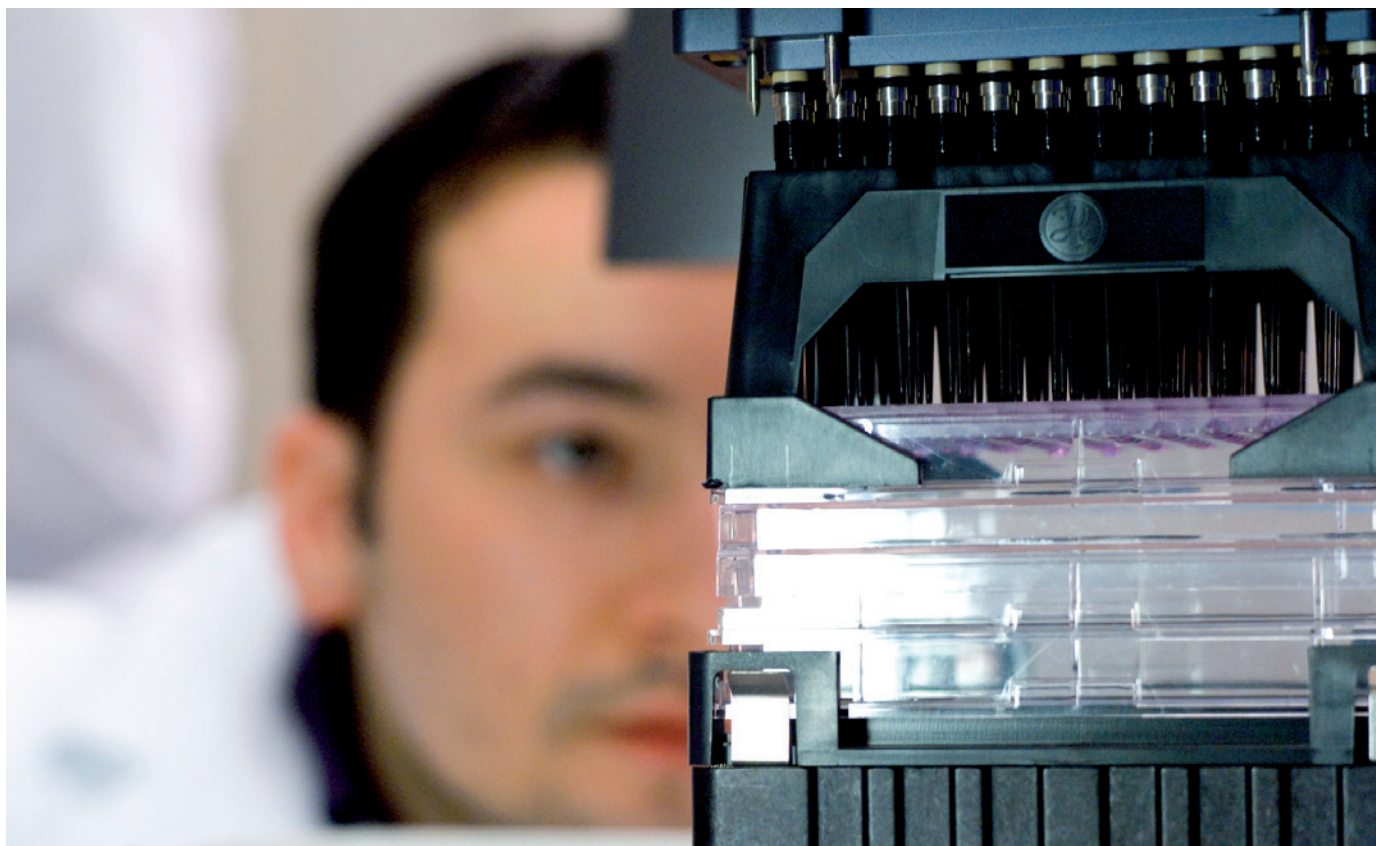
Nature **494**, 452–454 (2013); doi:10.1038/nature11914

In this Letter, the Centro de Astrobiología affiliation (number 11, associated with authors David Barrado and Jorge Lillo-Box) was listed incorrectly; the correct address is: 'Departamento Astrofísica, Centro de Astrobiología (INTA-CSIC), ESAC campus, PO Box 78, E-28691 Villanueva de la Cañada, Spain'. This has been corrected in the HTML and PDF versions of the manuscript.

TECHNOLOGY FEATURE

A BETTER BREW

Advances in cell culture media mean that scientists increasingly know what has gone into the mix, and cells are enjoying a more natural environment — even in the lab.



INSPIRO

Scientists explore three-dimensional environments in which they can foster the tissue-like growth of cell clusters.

BY VIVIEN MARX

Cells that thrive in the lab make for happy researchers. And vice versa: biology experiments can grind to a halt if investigators fail to get their cell cultures growing in the right nutrient medium.

That is why the market for such culture media is a lively one, with scores of commercial and home-brewed mixes available to help biologists to deal with all the different cell types that their experiments might require. But although the field of cell culturing can draw on generations of experience, making the right choice of medium is still more of an art than a science.

Even slight differences in media can have a large impact on cells — often for no clear reason. Many scientists mix their own culture media, but that can hamper the reproducibility of scientific findings. John Masters,

an experimental pathologist at University College London and editor of numerous books on animal and human cell culture, says that the recipe for such ‘home-brews’ can be difficult to follow owing to the sheer number of ingredients, as well as variations in purity and content between suppliers, variations between batches from a single supplier, and the difficulties of making relatively small quantities of a labile mixture of chemicals consistently.

However, as scientists come to terms with the importance of knowing exactly what their cells are thriving on, the field is becoming more rigorous. Some researchers, for example, are trying to eliminate culture-media components that originate from animals, because of fears that they could contaminate or infect potential human therapeutics down the line. Other investigators are trying to make growth media reproduce a natural

environment more realistically — for example by creating three-dimensional (3D) tissue structures.

SOME CELLS ARE HARD TO PLEASE

A prime example of the importance of good culture media is in the burgeoning field of induced pluripotent stem cells (iPSCs) — adult cells that have had their molecular clocks turned back to regain the any-fate-is-possible state of their infancy¹. These cells can be redirected to become many cell types, offering prospects for regenerative medicine² using lab-grown tissues to replace or renew aged, injured or diseased tissue in patients.

At the RIKEN Center for Developmental Biology in Kobe, Japan, for example, ophthalmologist Masayo Takahashi is hoping to gain approval soon for the first clinical trial of an iPSC-based treatment, for age-related macular degeneration, in which portions of the ►

► retina begin to die. Takahashi's goal is to replace the diseased parts of the retina with reprogrammed skin cells.

Meanwhile, basic researchers are exploring 'transdifferentiation': a genetic approach that converts one type of cell into a completely different one, skipping reversion to the stem-cell phase entirely. An example is the work of Rudolf Jaenisch and Yosef Buzanum at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts. Using a process based on cell culture, the researchers have shown that connective-tissue cells can be transformed into cells that express markers specific to Sertoli cells, which are normally found in the testis³. The results may help researchers to solve the puzzle of male infertility, and may pave the way for techniques for growing cell types that are currently difficult or even impossible to culture.

Buzanum says the "specific culture medium used is crucial for the particular fate the cells assume". Transdifferentiated cells, iPSCs or cultured neurons each need their own culture media tailored to the needs of the cells and cell type. The media might have or lack certain growth factors, for example, or create high or low oxygen levels, all of which allow cells to retain their normal characteristics and properties, says Buzanum.

More generally, stem-cell researchers in academic or industry labs need media and substrates to maintain and grow their cells, and to coax them down a series of developmental pathways, says Bradley Garcia, who directs technology and business development at Primorign Biosciences in Madison, Wisconsin, which develops and sells such products. Media tuned to cells' requirements can also keep differentiated stem cells, whether liver or heart cells or neurons, in culture for days, months or even more than a year.

Cells can be unpredictable, growing more readily in one medium than another for no apparent reason. And stem cells, according to Scott Monsma, senior director of research and development at Primorign, are "balanced on a razor-edge", and will differentiate in response, for example, to rough handling, overcrowding and stress. These factors make stem-cell-media development challenging, but at the same time, the medical potential of stem cells raises demand for such media, companies say.

LESS CAN BE MORE

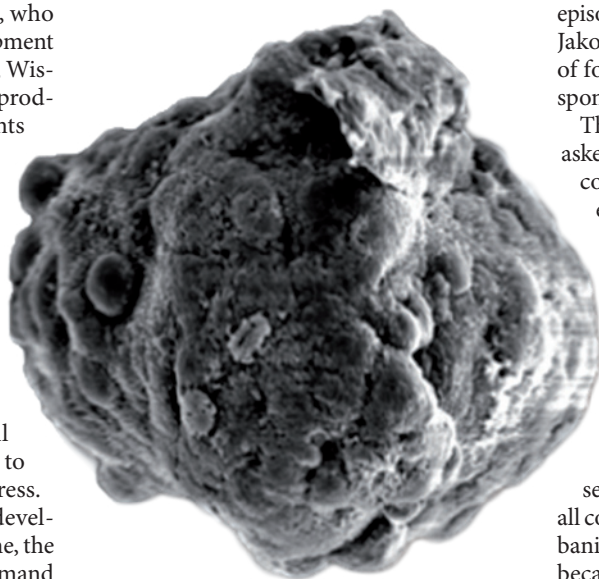
Some scientists use feeder-based systems to get their stem cells growing. In these systems, a layer of supporting cells, such as mouse embryonic fibroblasts, supply the medium with growth factors. But these systems can be prone to error, warns Erik Hadley, senior scientist in research and development at Stemcell Technologies, a spin-off from the British



Glauco Souza has made microtissues with *in vivo*-like morphology and protein production.

Columbia Cancer Agency that is based in Vancouver, Canada, and sells stem-cell media. Not only is each batch of feeder cells different, but it is also hard to control the amount and timing of excreted growth factors, making it difficult for researchers to know which ingredients make cells respond in what way.

To combat these issues, Stemcell Technologies sells a feeder-free stem-cell maintenance medium called mTeSR1. A follow-on product, TeSR2, is completely free of animal proteins,



An artistically enhanced scanning electron micrograph of a liver-cancer-cell spheroid.

and another, TeSR-E8, which was released in January, contains a set of eight components with a formulation based on work by James Thomson, a stem-cell researcher at the University of Wisconsin-Madison. The media are sold by Stemcell Technologies under patent licences from the university. Life Technologies in Carlsbad, California, also sells a version,

called Essential 8 Medium.

Defining exactly what has gone into a culture medium takes the field beyond alchemy and helps scientists to reproduce findings from colleagues, as well as to approach clinical applications, says Mikhail Kolonin, a stem-cell researcher at the University of Texas Health Science Center in Houston.

NO MORE ANIMAL PHARMA

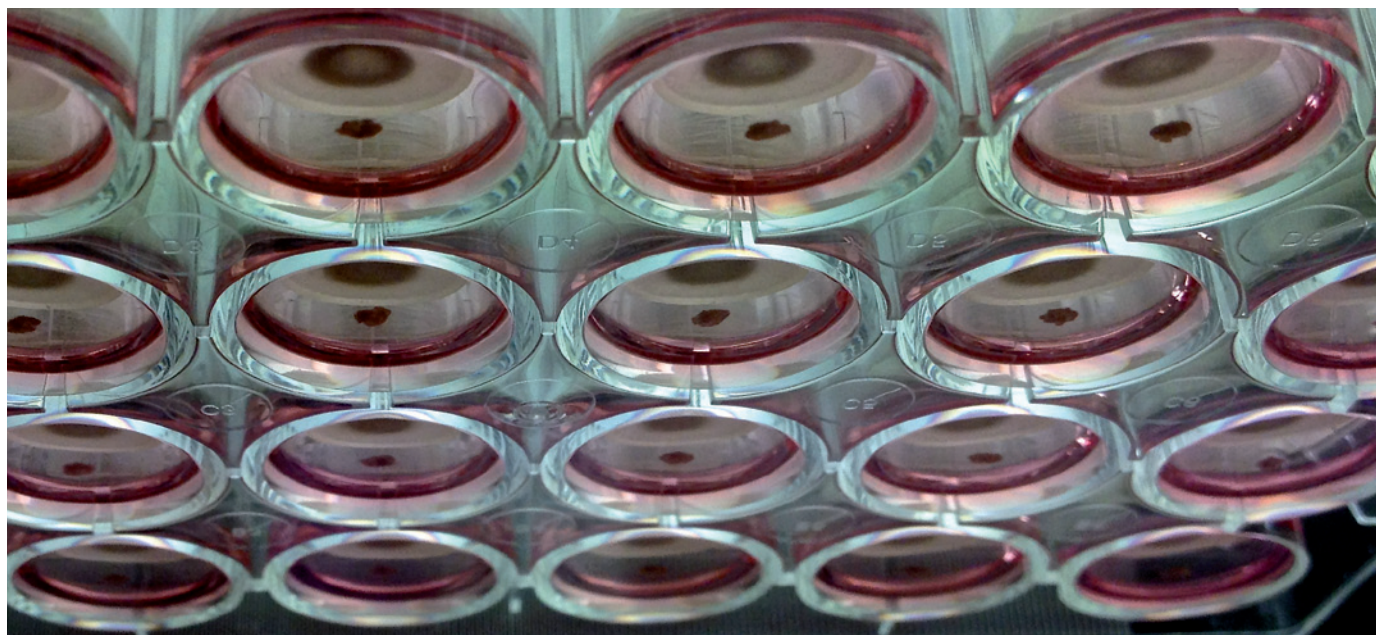
In addition to problems with feeder systems, another stumbling block to reproducibility is that the growth factors, proteins and other nutrients in stem-cell media typically come from fetal bovine serum, which can comprise up to one-fifth of a medium's volume, says Monsma. Each batch of serum — which is part of the blood — comes from a different animal and contains different amounts of components. "The point is, we don't know what's inside," says Kolonin. That is one reason why stem-cell scientists eyeing clinical applications are becoming wary of using products that contain serum.

Another reason is that cells grown in animal products for use in tissue transplantation can "potentially cause an immune response in patients", says Kolonin. Contamination can have even more serious consequences, as the experience with mad cow disease showed, says Nathan Allen, a product marketing manager in the cell-culture and bioprocessing business at Thermo Fisher Scientific, headquartered in Waltham, Massachusetts. In the mad-cow episode, a UK outbreak of variant Creutzfeldt-Jakob disease was caused by contamination of food with the infectious agent of bovine spongiform encephalopathy.

The US Food and Drug Administration has asked manufacturers to avoid animal-derived components in therapeutics. This preference affects preclinical research, because ideally, technology choices in the early stages of development should set the pattern for manufacturing further down the road, says Roberta Morris, business director at Thermo Fisher Scientific.

For that reason, new commercial media are increasingly serum-free, says Allen. His company offers a number of serum-free media and defined media free of all components originating from animals. But banishing animal products is not easy, if only because converting media containing serum or with undefined supplements into a more chemically defined version means massively reworking a proprietary recipe, which affects manufacturing.

All of this can make media expensive. Last autumn, Sigma-Aldrich in St Louis, Missouri, launched a stem-cell maintenance medium as part of its Stemline series. The medium is not completely free of animal products, but is composed of defined components and does not contain the types of crude protein ►



Technology by n3D Biosciences brings cells together in a dish using magnetic nanoparticles to levitate them in a magnetic field.

► preparation found in many formulations, such as serum or pituitary extracts, says Dan Allison, principal research scientist at Sigma-Aldrich. It was designed to cater for labs that are working towards industrial applications for stem cells and that will need high volumes of media, says the company.

Monsma says that creating media without animal components, using only chemical compounds and supplements of non-animal origin — such as human serum albumin or recombinant growth factors — means that the proteins must be expressed in human cells or bacteria, then purified and tested. His company and others are setting up capabilities to manufacture such media. For example, Primorigen is collaborating with several university labs to convert a differentiation medium into one that is animal-component-free.

For stem-cell researchers, shifting away from animal products means abandoning some lab staples, such as mouse feeder cells. Furthermore, some substances traditionally used to coat cell culture dishes are not animal-free. Matrigel — a product that was previously made by Becton, Dickinson of Franklin Lakes, New Jersey, but was sold to Corning Inc. in Corning, New York, last autumn — is a gel used to coat dishes, and is derived from a type of mouse tumour. Researchers at the University of Michigan in Ann Arbor have noted that, although Matrigel has helped scientists to define what iPSCs need, its animal origins and variability are problematic if cells are being cultured as eventual patient therapeutics⁴.

ENGINEERING IN THE MIX

Some scientists will be satisfied only by mixing their own cell culture media. They “tend to know what they are doing and be highly

experienced”, says Masters.

But most, he adds, “are generally not interested in the basics of how to do it properly, just the end product”. They want to be able to buy media off the shelf. Companies have begun catering for scientists who want more defined media. Firms interviewed by *Nature* say that their products are superior to home-brew because they can exert more quality control over the way they source, store, mix and evaluate ingredients, and can manufacture media under controlled conditions.

Engineered media can make a difference. For example, stem cells are deep frozen until their use in the lab, with scientists using a variety of cryopreservation media, including home-brews. But an ongoing challenge in the field is that most cells do not survive the thaw, says Hadley.

Towards the end of last year, Thermo Fisher Scientific began to sell a serum-free, animal-origin-free cryopreservation medium called HyCryo for standard cell lines, and a separate one, HyCryo-STEM, for stem cells. HyCryo-STEM is engineered to improve the recovery rate of stem cells after thawing. Scientists working with neural stem cells can usually recover only 10–20% of cells, and increasing that proportion is not easy with the typical home-brewed freezing media used in labs, says Cindy Neeley, a cell-culture specialist at Thermo Fisher. In tests, the company’s new medium is as good as home-brew, and for neural stem cells the recovery increased to 50–60%, she says.

Improving cell-culture environments also means improving containers.

“Most researchers are generally not interested in the basics of how to do it properly, just the end product.”

Taking an engineering approach, Po Ki Yuen, a bioengineer at Corning, has built a 96-well plate that nurtures growing cells while removing waste — which is toxic — and replenishing media, without an external pump⁵. Not only does the plate require less than the usual amount of daily media exchange, says Yuen — thus minimizing the need for human intervention and lowering the risk of contamination — but it also has fluid movement that is a bit more like that of the body than that of a classic lab vessel.

The idea for the plate, which emerged during a product-development session with two colleagues, he says, is to take advantage of pressure differences between wells that contain different amounts of fluid. Narrow strips of filter paper or a cellulose membrane connect the wells, so that fluid is forced to flow in a controllable way into the adjacent connected well until the liquid heights reach the same level in both. “The flow rate in our perfusion plate can be controlled by liquid height difference between connected wells, and the dimensions and pore size of the strip of cellulose membrane or filter paper,” says Yuen.

The 96-well version is not yet on the market but a 6-well version is, says Brian Douglass, a business-development manager at Corning. Cells in the 6-well version can last for at least 72 hours without media exchange, says Yuen. And, says Douglass, less-frequent media exchange means that “researchers get their weekends back”.

SPHERICAL THINKING

Scientists and companies are also exploring 3D environments in which to foster tissue-like growth of cell clusters, keeping cells close and in constant communication. In this kind of architecture, stem cells can grow into rounded

aggregates called embryoid bodies, which is part of the differentiation process.

This means that cells must not attach to the surface of their container, because if they do, they will grow in a spread-out single layer, says Neeley. Thermo Fisher Scientific has developed a series of dishes and multiple-well plates with a polystyrene surface that offers low-adhesion properties. Scaffolds can be used to shape cell clusters as they grow, but the three-dimensionality collapses once the scaffold is removed, like a tent without its supporting poles. They can also block a researcher's view through a microscope.

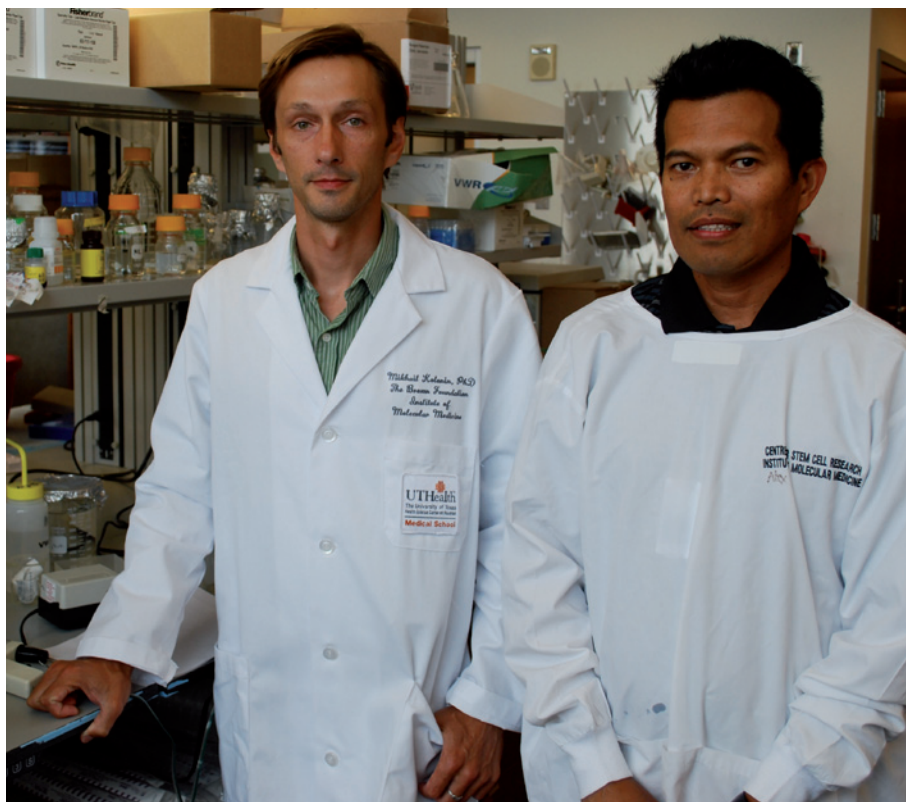
To cater for researchers seeking viable 3D cell-culture options, Thermo Fisher Scientific has developed a culture plate called Nunclon Sphera. "The cells, instead of sticking to the surface, aggregate with themselves and form a three-dimensional sphere in the culture environment," says Neeley. When cultured in this plate, cells grow into spheres that the scientists can transfer from one vessel to another using a pipette, without disrupting the form, she says. Customers are currently beta-testing the product.

Other plate-focused efforts rely on more radical architectural changes. Three-dimensional cell culture dates back more than 100 years, says Ross Harrison, a biologist at Johns Hopkins University in Baltimore, Maryland. He cultured neural tissue in a hanging drop of frog lymph and was able to observe live nerve cells sprouting axons, the long extensions through which neurons send messages to other neurons.

Now, a Swiss company, InSphero in Schlieren, is using the hanging-drop technique as the basis for a multiple-well plate made of the conventional polystyrene but with redesigned wells. After building a prototype, the Swiss Federal Institute of Technology (ETH) in Zurich set up the fledgling firm in the institute's technology park, says Jens Kelm, a biotechnologist formerly at the University of Zurich who founded InSphero four years ago along with University of Zurich colleague Wolfgang Moritz and ETH engineer Jan Lichtenberg. The firm is just moving into its own facilities.

Unlike typical round-bottomed wells, InSphero's wells have a V-shaped lower part, similar in appearance to a champagne flute. At its very bottom, the well is flat. In a hanging drop of medium, cells settle and grow as spheroids in a way that enables microscopy, says Kelm.

Getting the cells into the well also meant changing the well openings, which are shaped like a very narrow flower vase so that they fit tightly around a pipette tip. The researchers found that air-tight contact between the well opening and the pipette tip allowed them to deposit near-identical amounts into each well, which is important for making sure that the results are comparable between wells. To arrive



Mikhail Kolonin (left) and Alexes Daquinag at the University of Texas study adipose tissue and stem cells.

at the design, says Kelm, "we started experimenting with pipette tips, cut them off and put in drops from the top and looked at how they came out at the bottom".

In 2011, InSphero began a partnership with PerkinElmer in Waltham, Massachusetts, allowing the plates to be incorporated into PerkinElmer's automated screening instruments, which are used by drug companies. What began as a marketing deal has morphed into the companies collaborating on assay development; for example, they create plates that hold spheroids of liver microtissue ready for drug toxicity tests.

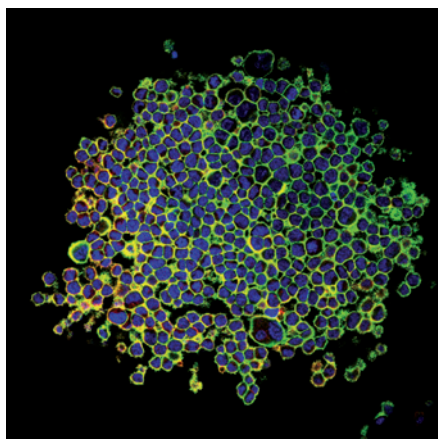
Kelm sees a broad international market for his technology. European laws that prohibit

the use of animals in cosmetics testing have left the industry clamouring for *in vitro* models, such as his microtissues. Drug developers and chemical companies also want cell-based assays to test toxicity. And hanging-drop technology can be used to culture stem cells, an area that could expand as these cells move towards medical applications.

Nadia Benkirane-Jessel, a biologist at the French National Institute of Health and Medical Research (INSERM) in Strasbourg, uses InSphero's technology to investigate ways to shorten recovery times for people undergoing bone-repair procedures and, potentially, bone regeneration. To position bone cells correctly for growth, Benkirane-Jessel seeds cells that have grown into spherical microtissues onto 3D nanofibres developed in her lab for use in mice. She also plans to use InSphero's technology for a product developed by her spin-off company, Artios Nanomed in Strasbourg, in the field of bone and cartilage regeneration.

CELL LEVITATION

Another academic spin-off that is developing 3D cell culture is n3D Biosciences in Houston. As chief scientific officer and company co-founder Glauco Souza explains, the technology seeds tissue by levitating cells and bringing them together⁶. The first step is to decorate cells with NanoShuttle, the company's magnetic nanoparticle assembly of gold and iron oxide crosslinked with polylysine, he says. Next, the cell culture dish is exposed to a magnetic field. "When the magnetic ►



Microtissues such as these colorectal-cancer cells can be grown in spheres without scaffolds.



Stem cells are often kept in liquid nitrogen at -196°C , but the recovery rate is low.

► field is applied, it brings the cells together while levitating them,” he says.

What keeps cells growing, Souza explains, is the cell–cell interaction that the levitation process promotes, which is more like the body’s environment than conventional cell culture. The technology also makes media exchange easier, because a magnet can hold the tissue in place, he says.

Souza, a physical chemist formerly at the MD Anderson Cancer Center in Houston, says that research with this technology at the company and in collaborating academic labs shows that the resultant microtissues have *in vivo*-like morphology and protein production, enabling them to be used in *in vitro* drug-testing models. n3D Biosciences has customers in university labs and pharmaceutical companies, and is focusing on high-throughput toxicity testing and drug development.

OVERNIGHT SUCCESS

Kolonin uses the technology to study fat tissue, and also considers it a possible environment for growing stem cells into organs. Recreating an organ in a dish requires all the organ’s cell types to be present and to make connections. In a flat dish, however, one cell type usually takes over because it happens to respond best to the media or to the plastic, and other cells are quickly lost, he says. That situation is different with the n3D technology. “You plate cells out, throw particles at them ... overnight, and put them into the magnetic field, and the next day you already have the spheroids,” he says. The spheroids include all the cell types. “It literally takes one day.”

Magnetic levitation has been a good way to model adipose tissue, Kolonin adds, and to culture stem cells while retaining their ability to differentiate⁷.

The magnetic particles may cause adverse consequences for the cells that contain them, but, these are a minority of the cells in culture. The microtissues stay together and the cells tend to spit out the particles, which then remain in the matrix outside the cells.

“There has been a boom of late in 3D formats, and I think the field is rapidly adopting and critically evaluating these technologies,” says Jeffrey Morgan, a bioengineer at Brown University in Providence, Rhode Island. He

“There has been a boom of late in 3D formats, and I think the field is rapidly adopting and critically evaluating these technologies.”

thinks that when cells contact, interact with and communicate with other cells rather than with artificial scaffolds, the cell culture more closely replicates the *in vivo* environment, especially that in solid organs such as the heart and liver, where cell density is high.

Morgan invented what he calls the 3D Petri dish, and in 2009 he founded a company: Microtissues, based in Providence. In a deal that went through last year, Sigma-Aldrich is distributing the dish. Morgan’s customers are academic biomedical researchers, pharmaceutical firms doing toxicity testing and cell-therapy companies exploring how to prepare cell clusters for possible transplantation.

The 3D Petri dish came about when, to guide the growth of tissue-like multicellular spheroids, Morgan and his graduate student Anthony Napolitano began making moulds in the lab. They wanted a material that was non-adhesive for cells, which is the “direct opposite” of the classic plastic Petri dish, says Morgan. At the same time, the researchers needed a

material that would not interfere with the small cell–cell adhesive forces that drive cell clustering. The material they chose was agarose, which forms a commonly used hydrogel and is made of 98% water, which is why Morgan says that his approach is “like sculpting water”.

A user casts molten agarose in the micro-moulds, allows it to gel, then removes the micro-moulded agarose and places it in a standard multiple-well dish. Cell media and cells are pipetted into the dish, and cells then settle by gravity into each of the micro-wells and self-assemble into a multicellular spheroid at the bottom of each moulded well. The micro-moulds can be autoclaved and reused to cast more gels.

The mould makes spheroids that are uniform in shape, says Morgan. Their size can be varied with the number of cells that are seeded. The cells are easier to harvest than in scaffold-based methods. They spill out when the gel is inverted, allowing further tests.

HOPES FLYING HIGH

Although stem-cell research and advances in cell culture are quickly advancing, viable cell therapies are years away from the market, says Chuck Oehler, chief executive of Primorigen. But companies such as his regularly get calls from people seeking stem-cell-based cures. Neither are stem-cell scientists immune to hope.

A researcher who did not wish to be identified is diabetic, and has been dependent on insulin for more than 30 years. A few years ago, he had a transplant of insulin-producing islet cells from cadavers, which allowed him to go for nearly one year without insulin injections and also lessened some of his symptoms, such as numbness in his fingers and toes.

“So the potential seems to be there, if the work we and others are doing to ensure production of cells with adequate, lasting function can be produced,” says the researcher. As a practitioner, he knows the scientific reality. “But given the impact that regenerative medicine can have on my quality of life and on my loved ones, it is easy to understand how those less familiar with the science and industry can be frustrated or impatient with the rate of progress.” ■

Vivien Marx is technology editor for *Nature* and *Nature Methods*.

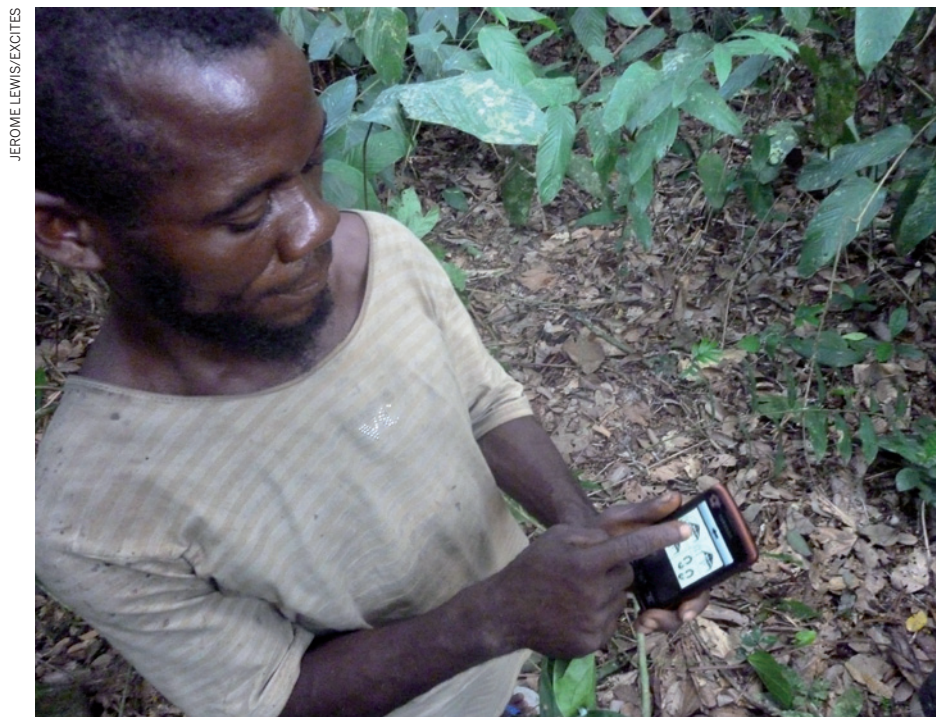
1. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
2. Robinson, D. A. & Daley, G. Q. *Nature* **481**, 295–305 (2012).
3. Buganim, Y. & Jaenisch, R. *Cell Cycle* **11**, 4485–4486 (2012).
4. Villa-Diaz, L. G., Ross, A. M., Lahann, J. & Krebsbach, P. H. *Stem Cells* **31**, 1–7 (2013).
5. Goral, V. N., Zhou, C., Lai, F. & Yuen, P. K. *Lab Chip* **13**, 1039–1043 (2013).
6. Souza, G. et al. *Nature Nanotechnol.* **5**, 291–296 (2010).
7. Daquinag, A. C., Souza, G. R. & Kolonin, M. G. *Tissue Eng. Part C* May, 336–344 (2013).

CAREERS

SEQUESTRATION Slashed budgets may put non-tenured US researchers at risk **p.261**

EUROPEAN UNION Proposed law would make researchers more mobile **p.261**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



Hunter-gatherers in the Congo Basin are using smartphones to track poaching and logging.

CITIZEN SCIENCE

Amateur experts

Involving members of the public can help science projects — but researchers should consider what they want to achieve.

BY TRISHA GURA

Equipped with smartphones, computers and do-it-yourself sampling kits, lay volunteers are tweeting about snowfall, questing for comets and measuring the microbes in their guts. They are part of a growing group of 'citizen scientists', networks of non-scientists who help to analyse or collect data as part of a researcher-led project. They learn about science and get a chance to participate, but the scientists involved stand to gain too. "There is huge amount of spare attention out there and a huge desire to do something real with it," says Christopher Lintott, an astronomer at the University of Oxford, UK, and chair of the Citizen Science Alliance, which hosts projects and advises researchers.

There are indications that in the past five years or so, citizen science has become more popular with researchers — and more likely to translate into a legitimate, publishable research project. It offers a means of doing substantial, thoughtful public outreach, and of tackling otherwise intractable, laborious or costly research problems.

But recruiting non-scientists comes with complications, including finding the right technical tools and partners to organize and execute projects with potentially thousands of data collectors. And scientists must ensure that the data are sound. Researchers thinking about entering the citizen-science fray should contact established associations for guidance and consider what they want from the end result — is it outreach alone, or something more?

It is difficult to measure the growth in citizen science accurately, in part because many ventures overlap with science-education efforts, but projects are definitely becoming more common. The Citizen Science Alliance began with one project in 2007 and has now hosted more than 20. The alliance has received 200 proposals for possible projects in the past year alone, and launched 10. A similar organization called SciStarter now boasts more than 450 projects. Citizen Science Central — co-founded by Rick Bonney, an ornithologist at Cornell University in Ithaca, New York, who in 1995 coined the term citizen science — lists 162 projects.

Some have given rise to peer-reviewed publications. Jake Weltzin, an ecologist at the University of Arizona in Tucson, is executive director of the USA National Phenology Network, which runs a project called Nature's Notebook. Citizen scientists track how climate affects the timing of life-cycle events in plants and animals. Weltzin's group jumped from one publication in 2007 — when citizen science elsewhere meant asking for access to people's computer-processing power to work with huge amounts of data — to more than 100 now. "This is a whole new way of doing science," says Weltzin. "Being able to think about and collect data at a continental scale."

STRENGTH IN NUMBERS

Citizen science can help researchers to address previously insoluble problems. For example, Justin Halberda, a psychologist at Johns Hopkins University in Baltimore, Maryland, wanted to study how cognition develops as people age. He originally thought that he would need to gather data from tens of thousands of people of all ages over the course of years, a feat that would be financially and logistically impossible for one research team with a limited budget.

So he reached out, asking volunteers to play a sort of video game that measures number sense — the ability to estimate how many items there are in a collection without actually counting them. The game involved watching yellow and blue dots flash briefly on a screen, then estimating whether there were more yellow or blue. Over the course of a few months, more than 13,000 individuals — ages 11 to 85 — took the test, and Halberda's team analysed the results.

The researchers not only built up a general description of people's number sense, but also unveiled a surprise: individuals don't achieve the best precision in number sense until ►

► they are about 30 years old, much older than most previous studies had suggested (J. Halberda *et al. Proc. Natl Acad. Sci. USA* **109**, 11116–11120; 2012). Before citizen science, “no scientist would have been able to generate these data”, says Halberda.

REACHING OUT

Citizen science can also educate and engage. Philip Brohan, a climate scientist at the Met Office Hadley Centre in Exeter, UK, and his colleagues build up data sets using weather records from nineteenth-century ships. The logs are handwritten, so they cannot be digitized automatically. The team set up an online project called Old Weather, in which citizen scientists pick a ship and read the logs’ accounts of epidemics, people falling overboard, getting stuck in ice and, of course, the weather. While following along, readers transcribe the information into online forms, which Brohan’s team analyses.

The project has been a huge success; readers get caught up in the accounts, as if they were diaries of the ship’s voyage. “People fall in love with the ships,” says Brohan. The weather has become just one part of a larger tale — but that part yields the science. Since the launch of Old Weather in October 2010, at least 16,400 citizen scientists have entered more than one million log pages to produce 1.6 million weather records, the basis for five atmospheric analyses.

Expanding on this kind of success, researchers are now using citizen science to improve their funding applications. Agencies often ask for some measure of how a research project will affect the lay public: the US National



Jack Meixner tracks juniper phenology in Texas as part of a citizen-science project on pollen.

Science Foundation (NSF), for example, requests a description of a project’s “broader impact” for many of its grants, including early-career development awards. Citizen science is a great way to meet your broader-impact requirements, says Weltzin.

It can also get the public involved in research that can inform environmental or governmental policy. In one case, citizens helped to collect data on noise levels near a scrapyard in the

London district of Deptford. Citizen scientists showed that the operation violated noise limits, and the UK Environment Agency revoked the scrapyard’s licence.

In perhaps the most avant-garde example of citizen science, members of communities in the Congo Basin are set to aid land and animal conservation. Jerome Lewis, a social anthropologist at University College London (UCL), was working with pygmy hunter-gatherers in Rwanda and other areas in the region when they told him about poachers killing animals and loggers destroying natural resources, such as sacred trees. Collaborating with Muki Haklay, a geographic information scientist at UCL, Lewis sought a way to help, and to document the damage for policy-makers. At first, it seemed that citizen science was not an option, because the hunter-gatherers could not read or write — let alone access the Internet.

But Haklay’s team designed smartphones with satellite-navigation systems and icons such as pictures of trees or traps. Community members carry the phones into the forests and press the relevant icon when standing in front of a landmark such as a sacred tree that could be targeted by loggers. The data streams back to Haklay’s team, which creates a map and report. These are shared with logging companies, which then decide whether to leave particular landmarks standing. A similar project to deal with poachers will launch this April. Data such as the locations of snares will be shared with policy-makers to attempt to reduce the killing of endangered animals. As an added feature, the phones come with a thermoelectric battery charger, which can convert heat from campfires to electricity. “If we find a way to go to the extremes of citizen science,” says Haklay, “we can do all kinds of really interesting stuff.”

GETTING STARTED

A call to arms

Launching a citizen-science project involves a few essential steps.

First, come up with a question that takes full advantage of a network of amateur data-collectors. It could be getting citizen scientists to track changes in rain acidity over time by measuring the weathering rate of marble gravestones in particular graveyards. The results could indicate changes in pollution and climate.

Next, seek guidance from people who have already built these kinds of projects, such as the Citizen Science Alliance (www.citizensciencealliance.org).

Network or correspond with individuals who are adept at marketing to educators and using social media to develop a project that is user friendly, easy to follow and attractive to many citizen-scientist recruits. Groups such as Citizen Science Central provide an online toolkit (go.nature.com/e1jjm5), and DataONE has a guide to data

management for citizen-science projects (go.nature.com/xfpdmh).

Invest in project design by, for example, creating a website, turning tasks into games, marketing and publicizing the project on social media and testing it on potential participants.

Plan to spend a lot of time and energy. It may take months or years to generate data for publication. Until then, the citizen-science effort may simply be another full-time responsibility.

Think about creative sources of funding, from private foundations to individual philanthropists or even corporations that want to support science education. Jason Osborne, co-founder of the citizen-science project Paleo Quest, reached out to teachers and then pitched an educational model for boosting interest in the sciences to corporate donors. “You can fund the research by funding the education,” he says. **T.G.**

WHAT TO WATCH FOR

Citizen science might appear to be all upside and little downside — a way to get others to do cheaply what researchers cannot or do not want to do. But that thinking can lead to trouble. “Citizen science is enormously expensive in terms of time and effort managing people, websites and databases,” says Brohan (see ‘A call to arms’).

People management is one of greatest challenges, says Andrea Wiggins, a social scientist who studies citizen science at the University of New Mexico in Albuquerque as part of DataONE, an NSF-funded project aimed at increasing the availability of Earth and environmental data. “You don’t necessarily know who is on the other end of a data point,” she says. It could be a retired botany professor reporting on wildflowers or a pure amateur with an untrained eye.

As a result, it is difficult to guarantee the quality of the data. Scientists have to design their projects and protocols for anyone to follow, and must perform regular quality control. Lintott and his team do exactly that as part of a project called Planet Hunters, in which citizen scientists sift through data from NASA’s Kepler spacecraft

to search for transit events — brief dips in the brightness of stars that occur when planets pass in front of them. To track how well citizen scientists can spot transits, and to measure the sensitivity of the system to different kinds, the Planet Hunters team inserts fake planet signals alongside the genuine data and measures how good people are at detecting them. In another quality-control measure, multiple users do the same task on the same data, so that mistakes are averaged out. And because users log in to use the system, the researchers can track which citizen scientists are best at which tasks and then give extra weight to results from the best performers.

The sheer scale of the projects can create quality-control challenges. The eBird project at Cornell uses citizen scientists to document the presence, absence and abundance of species across the United States. It receives 25 million observations a month, which are reviewed by a team of 500 volunteers, hand-picked for their experience. Each reviewer must sift through 4% of the observations to validate them, in essence looking at hundreds of thousands of observations. Project organizers are currently thinking about new ways to manage the big data sets, by either reviewing less or automating more.

Scientists also have to face the challenge of recruiting volunteers and keeping them engaged. “The science has to be romantic, in a way, so that people want to support the research behind it,” says Jason Osborne, president and co-founder of Paleo Quest, a citizen-science organization focused on palaeontology. Projects have to be interesting, tangible and involve discovery, he adds. In an offshoot of Paleo Quest called SharkFinder, for example, Osborne and co-founder Aaron Alford identify layers of rock riddled with shark fossils and take large samples from river swamps and other remote environments. The researchers then distribute the samples in kits to US classrooms. Any fossils that students discover are sent to the University of Maryland in College Park, where palaeontologists led by Bretton Kent verify the findings. The project currently samples fossil formations along the east coast of the United States, but Osborne hopes to expand his sampling work to Panama and other countries. “You put Panama in the kit, and kids are like, ‘Wow, I have a piece of Panama on my desk, and I am looking for fossil remains,’” says Osborne, noting that kids also love handling prehistoric fossils. “There has got to be that kind of wow factor.” And if they discover a new species, Osborne’s citizen scientists might be named on a publication — or even be given the opportunity to name the species.

“You are getting the information that you need at the same time that you are getting people involved.”

Whatever a volunteer’s motivation — even if it is just the joy of participating — scientists have to understand and nurture it. For Old Weather, Brohan’s team came up with a ranking system. Citizens receive the title of cadet when they join. After transcribing 30 pages of logs, they are promoted to lieutenant. The person who transcribes the most pages on a ship becomes the captain. Social media and online forums let the participants talk to each other.

Another way to keep volunteers engaged is to identify the best and invite them to do more. Scott Stevens, a research associate at the Cooperative Institute for Climate and Satellites in Asheville, North Carolina, works for a project in which citizen scientists classify images of storms. His team discovered that one-quarter of the data were coming from a handful of ‘power users’, who had classified more than 20 images each. In fact, a single citizen scientist contributed a full 7% of the activity on the website, more than the next five power users combined — and greater than all the least active half of the user base.

“Both these one-hit wonders and the super-users are important to us,” says Stevens, whose team wants to create something like Brohan’s ranking system to reward the best and most committed. As projects evolve, organizers can contact those super users and invite them to participate at a higher level, perhaps by helping to analyse data or to manage groups of other citizen scientists.

Finally, there is the challenge of getting citizen-science data through peer review. But that barrier is diminishing. “I have never come across a furious reaction against the idea of this,” says Lintott. “You just have to test your model as you would have you written a new computer code.”

Publications using data from citizen science are becoming more common, and even encouraged. Researchers at Princeton University in New Jersey, for example, have used data from Nature’s Notebook to expand a model of the timing of leaf-bud bursting from the Harvard Forest area in Massachusetts to the entire eastern seaboard of the United States. The team published its expanded model this year (S.-J. Jeong *et al. Geophys. Res. Lett.* **40**, 359–364; 2013). Not only did peer reviewers welcome the citizen-science data, but one actually gave advice on how to use the citizen-science model more effectively, says Weltzin.

If all goes well, citizen science is a way to communicate science, engage in outreach and accomplish research aims. “You are getting the information that you need at the same time that you are getting people involved,” says Weltzin. “It is like playing Whack-a-Mole with all hammers out. You meet all of your objectives at one time.” ■

Trisha Gura is a freelance writer in Boston, Massachusetts.

SEQUESTRATION

Staff cuts likely

US universities and research institutes will remain in stable financial health despite federal budget cuts, but staff reductions are likely, says a report by Moody’s Investors Service, an international credit-rating agency based in New York. In *The Sequester Series: Limited Impact on US Universities and Related Not-for-Profit Organizations*, published on 28 March, the agency notes that universities and research institutes will see an overall 5% cut in federal funds this year. The report is meant to reassure investors in the higher-education sector, but says that lay-offs, mainly of non-tenured researchers, could be needed to manage the cuts. Institutions are likely to increase their focus on interdisciplinary research and extramural collaborations to increase revenue and share costs, says Edith Behr, vice-president of Moody’s.

EUROPEAN UNION

Easier migration

Foreign researchers and students would find it easier to stay or work in the European Union (EU) under a proposed law. Legislation now in the draft phase would ensure that EU member states issue decisions on admission applications within 60 days; would permit researchers and students to move between states for 6–12 months or for the entire study period for Marie Curie or Erasmus Mundus fellows; and would let students and researchers stay in the EU for up to a year to find a job or start a business. The European Parliament and the Council of the EU are reviewing the proposal, which could be adopted by early 2016. Current regulations are “very fragmented”, says Michele Cercone, European Commission spokesman for home affairs. “This will fill a gap.”

MEDIA

Online journal club

A US medical journal is the first to use an online platform that lets participants discuss papers with their authors in real time. *Fertility and Sterility* has streamed two sessions of Journal Club Live on YouTube and will run a third in May, says platform developer Steven Palter, the journal’s new-media editor. Invited panellists take part in a video chat and YouTube viewers submit questions. Panellists and viewers for the first two sessions came from the United States, India and Spain. “This opens the discussion into a global collaboration,” says Palter.

SONIC ESTATE

A sound investment.

BY HENRY GEE

The distemper came away in skin-like flaps, dislodging gobbets of plaster and revealing the thin, rib-like slats of wood that made up the partition wall between two cramped attic rooms. Pausing from her labours, she brushed a strand of hair from her eyes and turned the stereo up full. Motörhead. *Ace of Spades*. No need to spare the batteries. Not any more.

He'd taken her to that rock disco. They'd just got their A-level results. The beads in her long hair flailed like maces, creating a space around them. University, then, and then graduate work. They'd parted, but when she set up the company, she brought him back in. Big mistake.

It wasn't long before she'd ripped the entire wall down. The air swirled with dust. Who knows what horrors it contained — fungal spores, shrapnel of horsehair, asbestos.

Early in her graduate work, searching around for something, she noticed how bubbles in superdense, superheated liquids bounced and popped to the metal on her lab stereo. Turn up the music past 11 and the cells gave out more energy than she put in. She was a lot quieter than the music, at least to begin with. Tuning frequencies, optimizing output and then, neutrons. Papers with cautious, noncommittal titles. Graduation. Patents pending. A start-up with her supervisor, who obligingly died, leaving her with the keys to the kingdom.

The house was an adversary, at least to begin with. It deceived in Escherish nooks that weren't in the plans, with windows giving out on to a clipped view nothing like the overgrown, half-wild surroundings she knew to be there. And that shrouded bedroom with the bat-eared gargoyles whose eyes followed her round the room. She could have sworn the horrible things screamed at her. That first-floor bedroom she chose for herself, at first, too huge, but which seemed to get much larger at night. And much colder. The first winter passed with her camped out in the library, close enough to the front door to make an escape should anything ever ...

He was the only person she knew who could handle the pressure as the company grew. Her skin

crawled remembering the day she phoned him. The catch in her voice, desperation, pleading. He was making big city bucks, marketing, management consulting, something she didn't grasp, but he had what she



hadn't. People skills. She just wanted to be R&D. She wasn't like him — *smooth*. And of course he had his price. He wanted to get smooth again like they were before, when, long haired, leather jacketed, they'd coupled, her screams, silenced in the pounding noise of the stereo. Hair cropped, after hours, across her desk, unguarded in corridors. He was so ... reptilian. But he wouldn't stop when she asked. Rougher, then, her memories, scrambled. Pain. The only way out was to sell the company. She fled beyond his grasp. Beyond anyone's grasp.

This lonely house on the coast amid scrub bent to a constant arctic breeze. Edwardian, ten bedrooms, vacant for years. She got contractors to fix the roof. Then, she decided, she'd be on her own. A life in labs had stripped her of any fear of night-shrieking plumbing or demoniac electrics.

A year passed, then two. She had all her groceries delivered, left at the doorstep. She didn't see another living soul. She broke the house, tamed it. It talked to her, cosseted her, it was all she needed. She smiled just once, when she bought — mail order, under

an assumed name — the fist-sized sonic fusion cell made by her old company. The generator that could power everything in her house for next to nothing, forever. The generator itself was submicroscopic. Most of the tennis-ball-sized, ceramic sphere was aerogel padding. Padding to absorb the noise, the incandescent rage generated by primal forces, trapped. Even now, when you put your ear to it, there was just the faintest hum. Like a rock band in heat, heard from a long way off. She wired it into the meter cupboard in the front hall.

She was wiping dust from her mouth when, through a sudden pause in the music, she heard a scrunch on the gravel outside. Panic. She switched off the stereo and suffocating silence flooded in on every side. The attic room had no window on the front, so she pounded down the back staircase two flights to the fanlight that looked out over the drive. It was just the gravel she heard: the sleek roadster made no noise. Fusion powered. Like everything else. The world she'd created. He was there. Him. No escape.

"Hey babe!" He looked up. Cocky, striped shirted. Leaning, self-assured, against the bonnet of the car. He'd seen her. "Knew you couldn't escape me forever. Don't be shy now, why don't we...?"

His next words were lost in a terrifying blast of noise. It started subsonic, but almost strong enough, she felt, to turn her inside out. Tiles began to shake from the roof. She felt rather than saw the front doors fly open beneath her, and, as the pitch rose, she knew that the hall cupboard had opened too, and the fusion generator had cracked open to the air. All the lights fused. All except for a blueish beam, directed out the front door, straight at him, straight at the car. She saw him, trapped, as the sonic battering broke both car and man into fragments, atomizing them. A last fizzle and a crack, and then primeval silence.

She stumbled down the stairs. The meter cupboard was as she'd left it, bolted. The fusion generator was entire, whole and silent. She walked out into the deserted front drive. She could have sworn she heard applause from the house. Or maybe it was just the gravel scrunching beneath her booted feet. ■

Henry Gee's *SF* trilogy *The Sigil* is published by ReAnimus press.

➔ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm